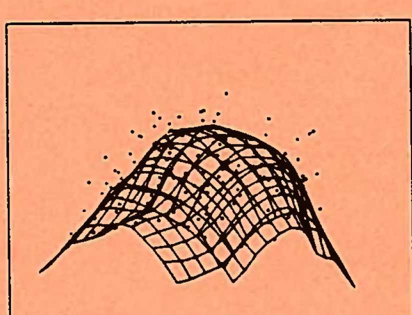
ON FREQUENCY POLYGONS AND AVERAGE SHIFTED HISTOGRAMS IN HIGHER DIMENSIONS

Nils Lid Hjort

Technical Report No. 22

February 1986

Laboratory for Computational Statistics



Department of Statistics
Stanford University

On Frequency Polygons and Average Shifted Histograms in Higher Dimensions

Nils Lid Hjort

Norwegian Computing Center

and

Stanford University

Abstract

Scott (1985a, 1985b) has recently studied two simple variations on the ordinary histogram, namely the frequency polygon and the average shifted histogram, and found that they are able to compete with for example kernel density estimators in performance while retaining the advantage of being conceptually and computationally simple. The present paper proposes a way of generalizing frequency polygons to d-dimensional space that performs better than Scott's generalization. Expressions for integrated mean squared error and for integrated mean absolute deviation plus integrated absolute bias are obtained for generalized frequency polygons, for average shifted histograms, and for generalized frequency polygons of average shifted histograms. These expressions are used to give guidelines for window sizes.

Key words and phrases: frequency polygons, average shifted histograms, multi-dimensional, integrated mean squared error, integrated mean absolute deviation, integrated absolute bias.

1. Introduction.

A simple way in which to smooth a (univariate) histogram is to connect midbin values with straight lines. This is the frequency polygon, which of course has been used for display purposes at least since 1900. It was not demonstrated until recently, however, that the gain of this simple linear smoothing is substantial, and that the frequency polygon comes a long way towards matching more sophisticated density estimators, while at the same time retaining the advantage of being conceptually and computationally simple; see Scott (1985a, 1985b).

It is not obvious how the notion of a frequency polygon should be extended to two and higher dimensions. Scott (1985a) gives one possible definition for the bivariate case, but a general d-dimensional definition along his lines would be awkward, and expressions for integrated mean squared error (IMSE), the traditional criterion by which to judge density estimators, would be very difficult to obtain. In Section 2 we discuss a natural extension termed the generalized frequency polygon, obtain the IMSE, and show that it performs better than Scott's version. We are also able to obtain an expression for another natural criterion, the integrated mean absolute deviation plus integrated absolute bias (IMAD + IAB), in the general d-dimensional case. These expressions provide guidelines for the choice of binwidths, and are informative for purposes of comparison with other density estimators.

Another neat construction of Scott (1985b) is the average shifted histogram, which shares with the frequency polygon the virtues of matching (for example) kernel estimators in performance while still being computationally more feasible when faced with the problem of evaluating the estimator many times from a large data set, which, for example, is the task of classification procedures built on nonparametric density estimation. Scott (1985b) obtains IMSE expressions for dimensions $d = 1, 2$. His results are supplemented in Section 3 with d-dimensional results for both IMSE and IMAD + IAB.

It is only natural to try out the two tricks mentioned above in tandem, and define the frequency polygon of the average shifted histogram. Again, Scott (1985b) has IMSE expressions for the uni- and bivariate case, but notes that explicit multivariate results are not generally available. Section 4 studies generalized frequency polygons of average shifted histograms, and once more expressions for IMSE and for IMAD + IAB are obtained.

Some consequences of these results are briefly discussed in Section 5, and comparisons with kernel density estimates are made.

Scott and others have emphasized the use of these histogram-type density estimates for display purposes, even in three and four dimensions (!), see for example Scott and Thompson (1983). Apart from general theoretical interest, the present work has been motivated by the possibility of using say generalized frequency polygons of average shifted histograms as building blocks in classifiers in symbol recognition and reconstruction of remotely sensed images. Typical characteristics of these technological problems are large training sets and high-dimensional feature vectors. The optimal classification rule depends upon a posteriori probabilities that again are expressed in terms of class densities, and a natural way to proceed is to estimate these nonparametrically. Classifiers built along these lines may use density estimators as "black boxes", and the need to display and scrutinize aspects of the data is secondary. These remarks also provide the motivation for having available a general d-dimensional theory.

To get started, let $X_1, \cdots, X_n$ be a sample of independent observations from an unknown smooth density f in $\mathbb{R}^d$. A common point of departure for later refinements will be a standard histogram density estimate $\hat{f}_0$ defined on a grid of cells with centres x_k and volume $h_1 \cdots h_d$, i.e.

$$\hat{f}_0(x) = (h_1 \cdots h_d)^{-1} Y_0(k)/n, \qquad x \in I_0(k) \tag{1.1}$$

where

$$I_0(k) = \prod_{j=1}^{d} \left(x_{k,j} - \frac{1}{2}h_j, \quad x_{k,j} + \frac{1}{2}h_j \right) \tag{1.2}$$

is cell number k, and $Y_0(k)$ is the number of X_i's falling in this cell.

Although exact expressions sometimes result from the reasoning in what follows, let us make clear that our analysis mainly is a large sample one, where $h_1 \to 0, \cdots, h_d \to 0$, $nh_1 \cdots h_d \to \infty$ as the number of observations tends to infinity. These requirements are standard and ensure that $\hat{f}_0$ above is consistent.

We shall not try to be as general as possible and shall be content to derive results for densities f with support equal to the union of the histogram cells and with continuous derivatives $\dot{f}_j(x) = \frac{\partial}{\partial x_j} f(x)$, $\ddot{f}_{j\ell}(x) = \frac{\partial^2}{\partial x_j \partial x_\ell} f(x)$, $\dddot{f}_{j\ell s}(x) = \frac{\partial^3}{\partial x_j \partial x_\ell \partial x_s} f(x)$ of first, second, and third order.

2. Generalized Frequency Polygons.

Consider the histogram defined in (1.1), (1.2). To generalize the notion of a frequency polygon, we should for a given x linearly combine nearby values of $\hat{f}_0$ and hopefully get an improved, smoothed version. Fix a particular k, and shift attention to the new cell

$$I(k) = \prod_{i=1}^{d}(x_{k,i}, x_{k,i} + h_i] = \prod_{i=1}^{d}(a_i - \frac{1}{2}h_i,\ a_i + \frac{1}{2}h_i], \tag{2.1}$$

where $a = x_k + \frac{1}{2}h$ is in the middle of 2^d histogram cells.

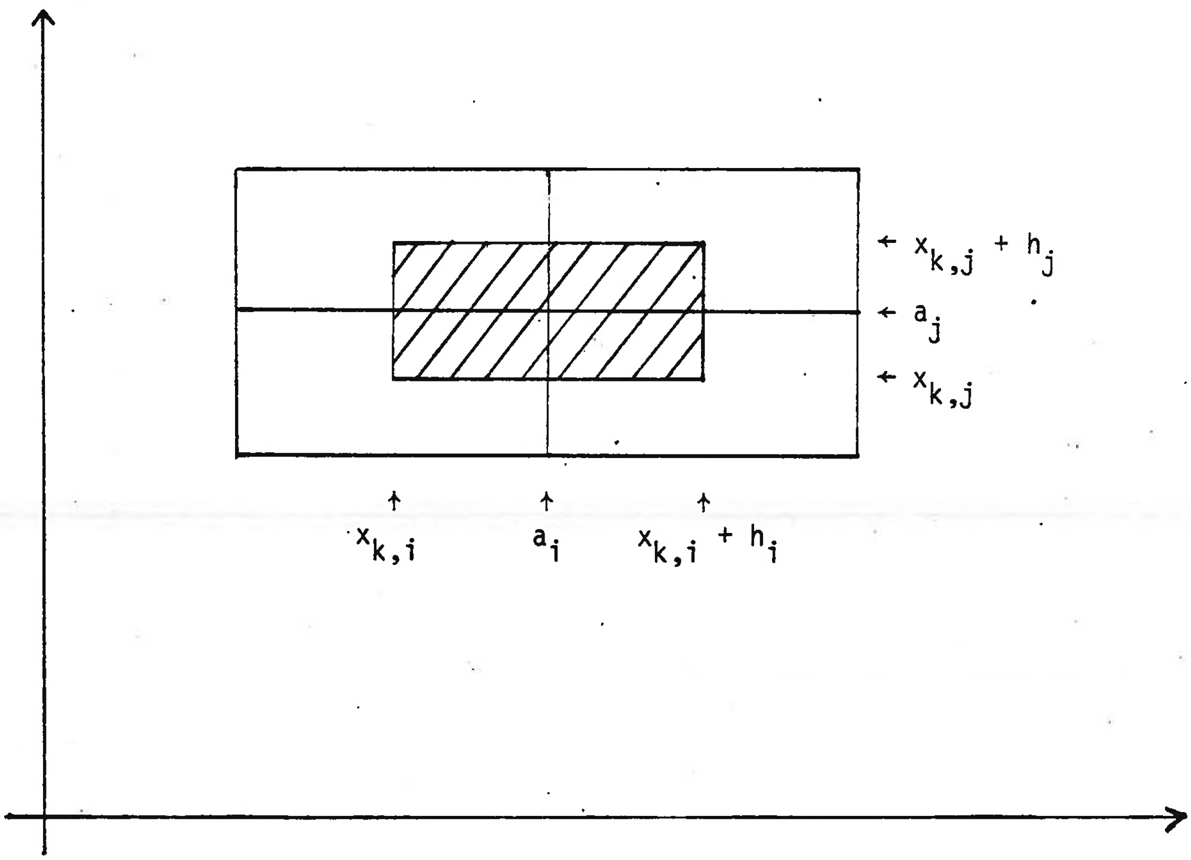


Figure 1. The GFP cell $I(k)$ lies within 2^d histogram cells $I_0(k; j_1, \cdots, j_d)$.

We want to define a generalized frequency polygon (GFP) in this "inner cell" by smoothing the 2^d histogram values $\hat{f}_0(x_{k,1}+j_1h_1, \cdots, x_{k,d}+j_dh_d) = \hat{f}_0(x_k+jh)$, $j = (j_1, \cdots, j_d) \in \{0,1\}^d$. Write

$$\hat{f}(x) = \sum_{j_1,\cdots,j_d} c_{j_1,\cdots,j_d}(x)(h_1\cdots h_d)^{-1}Y_0(k; j_1, \cdots, j_d)/n, \quad x \in I(k) \tag{2.2}$$

where $Y_0(k; j_1, \cdots, j_d) = Y_0(k; j)$ is the number of X_i's falling in

$$\begin{aligned} I_0(k;j) &= \prod_{i=1}^{d}(x_{k,i} + (j_i - \frac{1}{2})h_i,\ x_{k,i} + (j_i + \frac{1}{2})h_i] \\ &= \prod_{i=1}^{d}(a_i + (j_i - 1)h_i,\ a_i + j_i h_i]. \end{aligned} \tag{2.3}$$

The 2^d c_j-functions appearing in (2.2) are to be specified later. Natural immediate requirements are $c_j(x) \geq 0$, $\Sigma_j c_j(x) = 1$, making sure that $\hat{f}$ is a density in $I\!R^d$. The ordinary frequency polygon is of the form (2.2), with $d = 1$,

$$\begin{aligned} &c_0(x) = 1 - u(x), \quad c_1(x) = u(x); \\ &u(x) = (x - x_k)/h, \quad x \in [x_k, x_k + h] = [a - \frac{1}{2}h,\ a + \frac{1}{2}h]. \end{aligned} \tag{2.4}$$

2.1. Bias of the GFP.

One might consider estimates of the of the form (2.2) with weights $c_j(x)$ determined by the data; the discussion in the following is however limited to the case of non-random $c_j(x)$ functions.

The exact expectation of the GFP is

$$E\hat{f}(x) = \sum_j c_j(x)(h_1 \cdots h_d)^{-1} p_0(j), \quad x \in I(k)$$

where

$$p_0(j) = \int_{I_0(k;j)} f\,dx. \tag{2.5}$$

Approximations to $E\hat{f}(x)$ can now be worked out, for example based on Taylor expansions around x. It serves our present purpose best, however, to expand around the point $a = x_k + \frac{1}{2}h$.

We get

$$
\begin{aligned}
p_0(j)/(h_1\cdots h_d) &= \int_{I_0(k;j)} \Big\{ f(a) + \sum_{i=1}^{d} \dot{f}_i(a)(x_i - a_i) \\
&\quad + \frac{1}{2}\sum_{i,\ell} \ddot{f}_{i\ell}(a)(x_i - a_i)(x_\ell - a_\ell) + \cdots \Big\} dx/(h_1\cdots h_d) \\
&\doteq f(a) + \sum_{i=1}^{d} \dot{f}_i(a)\frac{1}{2}(-1)^{j_i+1}h_i + \sum_{i=1}^{d} \ddot{f}_{ii}(a)\frac{1}{6}h_i^2 \\
&\quad + \sum_{i\neq\ell} \ddot{f}_{i\ell}(a)\frac{1}{8}(-1)^{j_i+j_\ell}h_ih_\ell,
\end{aligned}
\tag{2.6}
$$

where $\doteq$ here and below is used after shaving off higher order terms. This makes exposition easier; regularity conditions are discussed later. Hence

$$
\begin{aligned}
E\hat{f}(x) &= \sum_j c_j(x)\Big\{ f(a) + \sum_{i=1}^{d} \dot{f}_i(a)\frac{1}{2}(-1)^{j_i+1}h_i + \sum_{i=1}^{d} \ddot{f}_{ii}(a)\frac{1}{6}h_i^2 \\
&\quad + \sum_{i\neq\ell} \ddot{f}_{i\ell}(a)\frac{1}{8}(-1)^{j_i+j_\ell}h_ih_\ell + \cdots \Big\} \\
&\doteq f(a) + \sum_{i=1}^{d} \dot{f}_i(a)\frac{1}{2}h_i\sum_j(-1)^{j_i+1}c_j(x)\,\cdot \\
&\quad + \sum_{i=1}^{d} \ddot{f}_{ii}(a)\frac{1}{6}h_i^2 + \sum_{i\neq\ell} \ddot{f}_{i\ell}(a)\frac{1}{8}h_ih_\ell\sum_j(-1)^{j_i+j_\ell}c_j(x).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathrm{bias}(x) &= E\hat{f}(x) - f(x) \\
&\doteq \sum_{i=1}^{d} \dot{f}_i(a)\Big\{\frac{1}{2}h_i\sum_j(-1)^{j_i+1}c_j(x) - (x_i - a_i)\Big\} \\
&\quad + \sum_{i=1}^{d} \ddot{f}_{ii}(a)\{\frac{1}{6}h_i^2 - \frac{1}{2}(x_i - a_i)^2\} \\
&\quad + \sum_{i\neq\ell} \ddot{f}_{i\ell}(a)\Big\{\frac{1}{8}h_ih_\ell\sum_j(-1)^{j_i+j_\ell}c_j(x) - \frac{1}{2}(x_i - a_i)(x_\ell - a_\ell)\Big\}.
\end{aligned}
\tag{2.7}
$$

Expressions for integrated squared bias and integrated absolute bias in terms of any given set of c_j-functions can now be obtained, but we will refrain from doing so until the best choice of c_j-functions has been settled on.

2.2. Variance of the GFP.

Using multinomial moments we get from (2.4)

$$
\begin{aligned}
\mathrm{Var}\{\hat{f}(x)\} &= (h_1\cdots h_d)^{-2}\sum_j c_j(x)^2\frac{1}{n}p_0(j)\{1-p_0(j)\} \\
&\quad -(h_1\cdots h_d)^{-2}\sum_{j\neq j'} c_j(x)c_{j'}(x)\frac{1}{n}p_0(j)p_0(j') \\
&= (nh_1\cdots h_d)^{-1}\sum_j c_j(x)^2 p_0(j)/(h_1\cdots h_d) \\
&\quad -\frac{1}{n}\Big\{\sum_j c_j(x)p_0(j)/(h_1\cdots h_d)\Big\}^2 \\
&\doteq (nh_1\cdots h_d)^{-1}f(a)\sum_j c_j(x)^2-\frac{1}{n}f(a)^2 \\
&\quad +\sum_{i=1}^{d}(nh_1\cdots h_d)^{-1}\dot{f}_i(a)\frac{1}{2}h_i\sum_j(-1)^{j_i+1}c_j(x)^2.
\end{aligned}
\tag{2.8}
$$

This approximation holds for x in $I(k)=(a-\frac{1}{2}h,\ a+\frac{1}{2}h]$.

2.3. The right choice: The linear blend interpolator.

Introduce

$$
u_i=u_i(x)=(x_i-x_{k,i})/h_i,\quad x_i\in[x_{k,i},x_{k,i}+h_i]=[a_i-\frac{1}{2}h_i,\ a_i+\frac{1}{2}h_i],
\tag{2.9}
$$

$i=1,\cdots,d$. u_i goes linearly from 0 to 1 as x_i moves from the left to the right side of the ith side of the cell $I(k)$, cf. (2.1). There is no loss of generality in representing the $c_j(x)$ functions in terms of $u_1,\cdots,u_d$.

Now turn to the choice of the 2^d $c_j(x)$ functions. In addition to $c_j(x)\geq 0$ and $\sum_j c_j(x)=1$ we should impose $\hat{f}=\hat{f}_0$ at the 2^d corners of $I(k)$, i.e. $c_{j_1,\cdots,j_d}(x)=1$ when $(u_1,\cdots,u_d)=(j_1,\cdots,j_d)$, that $\hat{f}$ is the plain average of the 2^d nearby corner values at the centre point a, and perhaps some symmetry. In some sense we want $c_j(x)$ to measure closeness of x to corner j.

Our choice is

$$
c_{j_1,\cdots,j_d}(x)=(1-u_1)^{1-j_1}u_1^{j_1}\cdots(1-u_d)^{1-j_d}u_d^{j_d}.
\tag{2.10}
$$

These functions satisfy the requirements above. More importantly, the algebraic expressions

(2.7), (2.8) for bias and variance simplify dramatically, and IMSE (strictly speaking, the leading terms of the IMSE) with any other choice of c_j-functions will be larger than IMSE with (2.10).

Now expressions for

$$\text{IMSE} = E\int\{\hat{f}(x)-f(x)\}^2dx = \int \text{Var}\{\hat{f}(x)\}dx + \int\{\text{bias}(x)\}^2dx \tag{2.11}$$

can be worked out. It is a matter of checking to arrive at

$$\sum_{j_1,\cdots,j_d}(-1)^{j_i+1}c_j(x) = 2u_i - 1 = 2\frac{x_i-a_i}{h_i}, \tag{2.12}$$

$$\sum_{j_1,\cdots,j_d}(-1)^{j_i+j_\ell}c_j(x) = (2u_i-1)(2u_\ell-1) = 4\frac{x_i-a_i}{h_i}\frac{x_\ell-a_\ell}{h_\ell}, \tag{2.13}$$

$$\sum_{j_1,\cdots,j_d}c_j(x)^2 = \prod_{i=1}^d\{(1-u_i)^2+u_i^2\}. \tag{2.14}$$

Hence from (2.7)

$$\text{bias}(x) \doteq \sum_{i=1}^d \ddot{f}_{ii}(a)\{\frac{1}{6}h_i^2 - \frac{1}{2}(x_i-a_i)^2\} \tag{2.15}$$

and

$$\begin{aligned}\int_{I(k)}\{\text{bias}(x)\}^2dx &\doteq \sum_{i=1}^d\{\ddot{f}_{ii}(a)\}^2\int_{I(k)}\{\frac{1}{6}h_i^2-\frac{1}{2}(x_i-a_i)^2\}^2dx \\ &\quad + \sum_{i\neq\ell}\ddot{f}_{ii}(a)\ddot{f}_{\ell\ell}(a)\int_{I_k}\{\frac{1}{6}h_i^2-\frac{1}{2}(x_i-a_i)^2\}\{\frac{1}{6}h_\ell^2-\frac{1}{2}(x_\ell-a_\ell)^2\}dx \\ &= \left[\sum_{i=1}^d\frac{49}{2880}h_i^4\{\ddot{f}_{ii}(a)\}^2 + \sum_{i\neq\ell}\frac{1}{64}h_i^2h_\ell^2\ddot{f}_{ii}(a)\ddot{f}_{\ell\ell}(a)\right]h_1\cdots h_d.\end{aligned}$$

Summing over all cells and approximating integrals with Riemann sums in the usual way we arrive at

$$\begin{aligned}\int\{\text{bias}(x)\}^2dx &\doteq \sum_{i=1}^d\frac{49}{2880}h_i^4\int(\ddot{f}_{ii})^2dx \\ &\quad + \sum_{i<\ell}\frac{1}{32}h_i^2h_\ell^2\int\ddot{f}_{ii}\ddot{f}_{\ell\ell}dx.\end{aligned} \tag{2.16}$$

Also, combining (2.8) and (2.14)

$$\int_{I(k)}\text{Var}\{\hat{f}(x)\}dx \doteq (nh_1\cdots h_d)^{-1}f(a)(\frac{2}{3})^dh_1\cdots h_d - \frac{1}{n}f(a)^2h_1\cdots h_d$$

so that

$$\int \operatorname{Var}\{\hat{f}(x)\}dx \doteq (\frac{2}{3})^d(nh_1\cdots h_d)^{-1} - \frac{1}{n}\int f^2 dx. \tag{2.17}$$

Hence an asymptotic IMSE approximation expression has been obtained; see Theorem 1 below.

Remark 1. The choice (2.10) was arrived at by the writer as a solution to the equations $\sum_j(-1)^{j_i+1}c_j(x) = 2(x_i - a_i)/h_i$, $\sum_j(-1)^{j_i+j_\ell}c_j(x) = 4(x_i - a_i)(x_\ell - a_\ell)/h_ih_\ell$, $i \neq \ell$, aiming at making the bias (2.7) have as small order as possible. He has later learned that in general, approximating a function in a rectangle by linear interpolation of its corner values, with weights as in (2.9), (2.10), is known in numerical analyst circles as *linear blend interpolation*. The GFP we propose is accordingly the *linear blend* of the ordinary histogram. (Note that the weights $c_{j_1,\cdots,j_d}(x)$ themselves are not linear in x.)

2.4. IMAD and IAB for the GFP.

We can in addition to IMSE study a criterion related to the L_1 distance $\int|\hat{f}-f|dx$, which in some ways is a more natural measure, see Devroye and Györfi (1985). The expected L_1 distance itself proves to be rather intractable mathematically, so we shall be content to study the natural and statistically meaningful upper bound

$$\begin{aligned} E\int\{|\hat{f}(x) - E\hat{f}(x)| + |E\hat{f}(x) - f(x)|\}dx \\ = \int[\operatorname{mad}\{\hat{f}(x)\} + |\operatorname{bias}(x)|]dx \\ = \text{IMAD} + \text{IAB}, \end{aligned} \tag{2.18}$$

writing $\operatorname{mad}\{\hat{f}(x)\} = E|\hat{f}(x) - E\hat{f}(x)|$ for the mean absolute deviation and $|\operatorname{bias}(x)|$ for the absolute bias. Note the similarity of (2.18) to the traditional criterion (2.11).

Sometimes we shall take interest in IMAD and IAB evaluated over some bounded region instead of over all of $\mathbb{R}^d$.

We shall in fact sometimes only give upper bounds for $\int|\operatorname{bias}(x)|dx$, since exact calculations tend to be difficult (but possible, as opposed to the exact expected L_1 distance, which borders on the impossible), and since the resulting expressions do not convey as useful information as the upper bounds. For illustration of this point, consider $\int_{I(k)}|\sum_{i=1}^d b_i(x_i - a_i)|dx$. This integral may be explicitly evaluated in terms of $b_1,\cdots,b_d$ and the widths $h_1,\cdots,h_d$, but the answer is less informative and useful than the simple upper bound $\sum_{i=1}^d \frac{1}{4}|b_i|h_i(h_1\cdots h_d)$.

Now consider IMAD and IAB for the chosen GFP. From (2.15) we get

$$
\begin{aligned}
\int_{I(k)} |\text{bias}(x)|dx &\doteq \int_{I(k)} |\sum_{i=1}^{d} \ddot{f}_{ii}(a)\{\frac{1}{6}h_i^2 - \frac{1}{2}(x_i - a_i)^2\}|dx \\
&\leq \sum_{i=1}^{d} |\ddot{f}_{ii}(a)| \int_{[a-\frac{1}{2}h,\, a+\frac{1}{2}h]} \{\frac{1}{6}h_i^2 - \frac{1}{2}(x_i - a_i)^2\}dx \\
&= \sum_{i=1}^{d} \frac{1}{8}|\ddot{f}_{ii}(a)|h_i^2(h_1 \cdots h_d). \qquad (2.19)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\hat{f}(x) - E\hat{f}(x) &= \sum_j c_j(x)(h_1 \cdots h_d)^{-1}\{\frac{1}{n}Y_0(k;j) - p_0(j)\} \\
&= \sum_j c_j(x)(h_1 \cdots h_d)^{-1} N_0(j)[p_0(j)\{1 - p_0(j)\}/n]^{1/2} \\
&= f(a)^{1/2}(nh_1 \cdots h_d)^{-1/2} \sum_j c_j(x) N_0(j) \left[\frac{p_0(j)\{1 - p_0(j)\}}{f(a)h_1 \cdots h_d}\right]^{1/2},
\end{aligned}
$$

writing

$$
N_0(j) = N_0(j_1, \cdots, j_d) = \frac{Y_0(k;j) - np_0(j)}{[np_0(j)\{1 - p_0(j)\}]^{1/2}}. \qquad (2.20)
$$

These variables are asymptotically independent and standard normally distributed by a triangular and multivariate version of the Lindeberg theorem, as long as $nh_1 \cdots h_d \to \infty$. (Even though $p_0(j) \to 0$ by (2.6) one still has $np_0(j) \to \infty$.) They are also uniformly integrable, since $EN_0(j)^2 \equiv 1$. Hence

$$
E|\sum_j c_j(x)N_0(j)| \doteq E|N(0, \sum_j c_j(x)^2)| = (\frac{2}{\pi})^{1/2}\{\sum_j c_j(x)^2\}^{1/2}.
$$

This suggests

$$
\begin{aligned}
\int_{I(k)} \text{mad}\{\hat{f}(x)\}dx &\doteq f(a)^{1/2}(nh_1 \cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2} \int_{I(k)} \{\sum_j c_j(x)^2\}^{1/2}dx \\
&= f(a)^{1/2}(nh_1 \cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2} \prod_{i=1}^{d} \int_{a_i - h_i/2}^{a_i + h_i/2} [\{1 - u_i(x)\}^2 + u_i(x)^2]^{1/2}dx_i \\
&= f(a)^{1/2}(nh_1 \cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2} \left\{\frac{1}{2} + \frac{1}{2}\frac{\log(1 + \sqrt{2})}{\sqrt{2}}\right\}^d h_1 \cdots h_d
\end{aligned}
$$

and a corresponding IMAD expression.

For fixed $n, h_1, \cdots, h_d$ there will always be many cells left with $np_0(j) \doteq f(a)\; nh_1 \cdots h_d$ small, even if $nh_1 \cdots h_d$ is large, so we cannot expect for example $E|N_0(j)| \doteq (\frac{2}{\pi})^{1/2}$ to be a good approximation for *all* cells. (In fact, about the best one can get is $\left|E|N_0(j)| - (\frac{2}{\pi})^{1/2}\right| \le B/[np_0(j)\{1 - p_0(j)\}]^{1/2}$ for some constant B; see the Appendix.) The unavailability of a closed form expression for $E|\hat{f}(x) - E\hat{f}(x)|$, therefore, will lead to good approximations for IMAD only if f has bounded support, or if $\mathrm{mad}\{\hat{f}(x)\}$ is integrated over a bounded region only.

Theorem 1. Let the density f in $I\!R^d$ have continuous derivatives $\dot{f}_i$, $\ddot{f}_{ij}$, $\dddot{f}_{ijk}$, and let $h_1 \to 0, \cdots, h_d \to 0$, but $nh_1 \cdots h_d \to \infty$. Then for the generalized frequency polygon defined in (2.2), (2.10):

$$\begin{aligned}
\mathrm{IMSE} &= \int \mathrm{Var}\{\hat{f}(x)\}dx + \int \{\mathrm{bias}(x)\}^2 dx \\
&= (\frac{2}{3})^d (nh_1 \cdots h_d)^{-1} - \frac{1}{n}\int f^2 dx + \sum_{i=1}^{d} \frac{49}{2880} h_i^4 \int (\ddot{f}_{ii})^2 dx \\
&\quad + \sum_{i<\ell} \frac{1}{32} h_i^2 h_\ell^2 \int \ddot{f}_{ii}\ddot{f}_{\ell\ell} dx + O(\sum_{i=1}^{d} h_i^5 + \sum_{i=1}^{d} h_i^2/(nh_1 \cdots h_d));
\end{aligned}$$

$$\begin{aligned}
\mathrm{IMAD} + \mathrm{IAB} &= \int \mathrm{mad}\{\hat{f}(x)\}dx + \int |\mathrm{bias}(x)| dx \\
&\le (\frac{2}{\pi})^{1/2}\{1 + \frac{1}{\sqrt{2}}\log(1+\sqrt{2})\}^d (\frac{1}{2})^d \int f^{1/2} dx\, (nh_1 \cdots h_d)^{-1/2} \\
&\quad + \sum_{i=1}^{d} \frac{1}{8} h_i^2 \int |\ddot{f}_{ii}| dx + O((nh_1 \cdots h_d)^{-1} + \sum_{i=1}^{d} h_i^2 (nh_1 \cdots h_d)^{-1/2} + \sum_{i=1}^{d} h_i^3).
\end{aligned}$$

The IMSE expression needs to have $\int f^2\, dx$, $\int (\dot{f}_i)^2 dx$, $\int (\ddot{f}_{ij})^2 dx$, $\int (\dddot{f}_{ijk})^2 dx$ finite. The IMAD + IAB expression holds provided the integrals are evaluated over some fixed *bounded region* contained in the interior of the support of f.

Some details pertaining to the proof of this theorem are given in the Appendix.

Remark 2. The IMSE expression here is better than the one obtained by Scott (1985a, equation (7.1)) for the case $d = 2$, since he used a less efficient choice of functions $c_{0,0}(x)$, $c_{0,1}(x)$, $c_{1,0}(x)$, $c_{1,1}(x)$ than the linear blend weights (2.11).

Remark 3. If the support of f is $I\!R^d$ then the IMAD + IAB expression holds over each bounded region. (Actually, taking resort to bounded regions is only necessary for IMAD,

not for IAB.) If f has compact support (but possibly defined only on its interior) then the expression holds provided only that the functions $|\dot{f}_i \dot{f}_j|/f^{3/2}$ and $|\ddot{f}_{ij}|/f^{1/2}$ have finite integrals.

3. Average Shifted Histograms

Consider again a histogram density estimate $\hat{f}_0$ of the form (2.1) in $\mathbb{R}^d$. Choose integers $m_1, \cdots, m_d$ and consider the smaller binwidths

$$\delta_j = h_j/m_j, \quad j = 1, \cdots, d. \tag{3.1}$$

$m_1 \cdots m_d$ new histograms can be constructed by moving the grid of cells an amount $i_j \delta_j$, $j = 1, \cdots, d$; $i_j = 0, 1, \cdots, m_j - 1$. Scott (1985a) proposes taking the average of these shifted histograms, i.e.

$$\hat{f}_{\text{ASH}}(x) = \frac{1}{m_1 \cdots m_d} \sum_{i_1=0}^{m_1-1} \cdots \sum_{i_d=0}^{m_d-1} \hat{f}_{0,\,\text{shifted}\; i_1\delta_1, \cdots, i_m\delta_m}(x).$$

$\hat{f}_{\text{ASH}}$ is constant on each of the many smaller cells of volume $\delta_1 \cdots \delta_d$. Single out one of these, say $C_0 = (x_0 - \frac{1}{2}\delta,\ x_0 + \frac{1}{2}\delta] = \prod_{j=1}^d (x_{0,j} - \frac{1}{2}\delta_j,\ x_{0,j} + \frac{1}{2}\delta_j]$, and write $Y(i) = Y(i_1, \cdots, i_d)$ for the number of X_j's that fall in the cell

$$\begin{aligned} C_0(i) &= (x_0 + (i - \frac{1}{2})\delta,\ x_0 + (i + \frac{1}{2})\delta] \\ &= \prod_{j=1}^{d} (x_{0,j} + (i_j - \frac{1}{2})\delta_j,\ x_{0,j} + (i_j + \frac{1}{2})\delta_j]. \end{aligned} \tag{3.2}$$

Then

$$\hat{f}_{\text{ASH}}(x) = (h_1 \cdots h_d)^{-1} \sum_{i_1=1-m_1}^{m_1-1} \cdots \sum_{i_d=1-m_d}^{m_d-1} \left(1 - \frac{|i_1|}{m_1}\right) \cdots \left(1 - \frac{|i_d|}{m_d}\right) Y(i_1, \cdots, i_d)/n \tag{3.3}$$

for x in the particular cell C_0; see Scott (1985b, Section 2).

Scott (op. cit.) finds IMSE expressions for $d = 1, 2$, but notes that explicit multivariate IMSE results are not generally available. His results are complemented below with such d-dimensional results. The present treatment will differ only mildly from his.

3.1. Bias of the ASH.

Write

$$w(i) = w(i_1, \cdots, i_d) = \left(1 - \frac{|i_1|}{m_1}\right) \cdots \left(1 - \frac{|i_d|}{m_d}\right) \tag{3.4}$$

and let $p(i) = p(i_1, \cdots, i_d) = \int_{C_0(i)} f dx$. Then from (3.3)

$$E\hat{f}_{\mathrm{ASH}}(x) = (h_1 \cdots h_d)^{-1} \sum_{\substack{1-m_j \le i_j \le m_j - 1 \\ j=1,\cdots,d}} w(i_1, \cdots, i_d) p(i_1, \cdots, i_d),$$

i.e. the exact expectation involves f-probabilities for $(2m_1 - 1) \cdots (2m_d - 1)$ cells. They are all contained in a cell with volume $2h_1 \cdots 2h_d$ around x_0, however, so approximations to $p(i_1, \cdots, i_d)$ based on Taylor expansion of f around x_0 may still be accurate enough. We get

$$\begin{aligned} p(i_1, \cdots, i_d)/(\delta_1 \cdots \delta_d) &= \int_{(x_0 + (i - \frac{1}{2})\delta, x_0 + (i + \frac{1}{2})\delta]} f(x) dx / (\delta_1 \cdots \delta_d) \\ &\doteq f(x_0) + \sum_{j=1}^{d} \dot{f}_j(x_0) i_j \delta_j + \frac{1}{2} \sum_{j=1}^{d} \ddot{f}_{jj}(x_0)(i_j^2 + \frac{1}{12})\delta_j^2 \\ &+ \frac{1}{2} \sum_{j \neq \ell} \ddot{f}_{j\ell}(x_0) i_j i_\ell \delta_j \delta_\ell + \frac{1}{6} \sum_{j=1}^{d} \dddot{f}_{jjj}(x_0)(i_j^3 + \frac{1}{4} i_j)\delta_j^3 \\ &+ \frac{1}{6} 3 \sum_{j \neq \ell} \dddot{f}_{jj\ell}(x_0)(i_j^2 + \frac{1}{12}) i_\ell \delta_j^2 \delta_\ell \\ &+ \frac{1}{6} \sum_{j,\ell,s \text{ distinct}} \dddot{f}_{j\ell s}(x_0) i_j i_\ell i_s \delta_j \delta_\ell \delta_s. \end{aligned} \tag{3.5}$$

Notice for the following that $w(i)/(m_1 \cdots m_d) = \prod_{j=1}^{d} \frac{1}{m_j}\left(1 - \frac{|i_j|}{m_j}\right)$ defines a probability distribution for $(i_1, \cdots, i_d)$ over $\prod_{j=1}^{d}\{1 - m_j, \cdots, 0, \cdots, m_j - 1\}$, with $i_1, \cdots, i_d$ independent, with odd moments equal to zero, and with $\mathcal{E}(i_j)^2 = \frac{1}{6}(m_j^2 - 1)$. Using this we obtain, for $x \in C_0$,

$$\begin{aligned} E\hat{f}_{\mathrm{ASH}}(x) &= \sum_{i_1,\cdots,i_d} \frac{w(i)}{m_1 \cdots m_d} \frac{p(i)}{\delta_1 \cdots \delta_d} \\ &= f(x_0) + \frac{1}{2} \sum_{j=1}^{d} \ddot{f}_{jj}(x_0)\{\frac{1}{6}(m_j^2 - 1) + \frac{1}{12}\}\delta_j^2 + O\left(\sum_{j=1}^{d} \delta_j^4\right). \end{aligned} \tag{3.6}$$

After subtracting a similar expression for $f(x)$ we obtain an approximation for the bias:

$$
\begin{aligned}
E\hat{f}_{\text{ASH}}(x) - f(x) \doteq &-\sum_{j=1}^{d} \dot{f}_j(x_0)(x_j - x_{0,j}) \\
&+\sum_{j=1}^{d} \ddot{f}_{jj}(x_0)\{(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 - \frac{1}{2}(x_j - x_{0,j})^2\} \\
&-\frac{1}{2}\sum_{j\neq\ell} \ddot{f}_{j\ell}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell}) \\
&-\frac{1}{6}\sum_{j,\ell,s} \dddot{f}_{j\ell s}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})(x_s - x_{0,s}).
\end{aligned}
\tag{3.7}
$$

Hence

$$
\begin{aligned}
\{\text{bias}(x)\}^2 \doteq &\left\{\sum_{j=1}^{d} \dot{f}_j(x_0)(x_j - x_{0,j})\right\}^2 \\
&+\left[\sum_{j=1}^{d} \ddot{f}_{jj}(x_0)\{(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 - \frac{1}{2}(x_j - x_{0,j})^2\}\right]^2 \\
&+\frac{1}{4}\left\{\sum_{j\neq\ell} \ddot{f}_{j\ell}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})\right\}^2 \\
&-2\left\{\sum_{j=1}^{d} \dot{f}_j(x_0)(x_j - x_{0,j})\right\}\left[\sum_{j=1}^{d} \ddot{f}_{jj}(x_0\{(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 - \frac{1}{2}(x_j - x_{0,j})^2\}\right] \\
&+\left\{\sum_{j=1}^{d} \dot{f}_j(x_0)(x_j - x_{0,j})\right\}\left\{\sum_{j\neq\ell} \ddot{f}_{j\ell}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})\right\} \\
&-\left[\sum_{j=1}^{d} \ddot{f}_{jj}(x_0)\{(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 - \frac{1}{2}(x_j - x_{0,j})^2\}\right]\left\{\sum_{j\neq\ell} \ddot{f}_{j\ell}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})\right\} \\
&+\frac{1}{3}\left\{\sum_{j=1}^{d} \dot{f}_j(x_0)(x_j - x_{0,j})\right\}\left\{\sum_{j,\ell,s} \dddot{f}_{j\ell s}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})(x_s - x_{0,s})\right\} \\
&-\frac{1}{3}\left\{\sum_{j=1}^{d} \ddot{f}_{jj}(x_0)(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2\right\}\left\{\sum_{j,\ell,s} \dddot{f}_{j\ell s}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})(x_s - x_{0,s})\right\},
\end{aligned}
$$

shaving off higher order terms.

Next integrate this over $C_0 = (x_0 - \frac{1}{2}\delta,\ x_0 + \frac{1}{2}\delta]$. Many details later one gets

$$\begin{aligned}\int_{C_0} \{\text{bias}(x)\}^2 dx \doteq & \Big[\sum_{j=1}^d \dot{f}_j(x_0)^2 \frac{1}{12}\delta_j^2 \\ & + \sum_{j=1}^d \ddot{f}_{jj}(x_0)^2 \frac{1}{144}\delta_j^4 \{(m_j^2 - \frac{1}{2})^2 - (m_j^2 - \frac{1}{2}) + \frac{144}{320}\} \\ & + \sum_{j\neq \ell} \ddot{f}_{jj}(x_0)\ddot{f}_{\ell\ell}(x_0)\frac{1}{144}\delta_j^2\delta_\ell^2(m_j^2-1)(m_\ell^2-1) \\ & + \frac{1}{2}\sum_{j\neq\ell}\ddot{f}_{j\ell}(x_0)^2\frac{1}{144}\delta_j^2\delta_\ell^2 + \sum_{j=1}^d \dot{f}_j(x_0)\,\dddot{f}_{jjj}(x_0)\frac{1}{240}\delta_j^4 \\ & + \sum_{j\neq\ell}\dot{f}_j(x_0)\,\dddot{f}_{j\ell\ell}(x_0)\frac{1}{144}\delta_j^2\delta_\ell^2\Big]\delta_1\cdots\delta_d.\end{aligned}$$

Summing up over all cells and using $\delta_j = h_j/m_j$ we arrive at

$$\begin{aligned}\int_{I\!R^d}\{\text{bias}(x)\}^2\, dx \doteq & \sum_{j=1}^d \frac{1}{12}\frac{h_j^2}{m_j^2}\int(\dot{f}_j)^2 dx \\ + \sum_{j=1}^d h_j^4 & \Big[\frac{1}{144}\{(1-\frac{1}{2m_j^2})(1-\frac{3}{2m_j^2}) + \frac{144}{320}\frac{1}{m_j^4}\}\int(\ddot{f}_{jj})^2dx + \frac{1}{240}\frac{1}{m_j^4}\int \dot{f}_j\,\dddot{f}_{jjj}\,dx\Big] \\ + \sum_{j\neq\ell} h_j^2h_\ell^2\frac{1}{144} & \Big[(1-\frac{1}{m_j^2})(1-\frac{1}{m_\ell^2})\int\ddot{f}_{jj}\ddot{f}_{\ell\ell}dx \\ & + \frac{1}{2}\frac{1}{m_j^2}\frac{1}{m_\ell^2}\int(\ddot{f}_{j\ell})^2dx + \frac{1}{m_j^2}\frac{1}{m_\ell^2}\int \dot{f}_j\,\dddot{f}_{j\ell\ell}\,dx\Big].\end{aligned} \tag{3.8}$$

Somewhat hidden here is the surprising fact that $\frac{1}{12}\delta_j^2\sum_k \dot{f}_j(x_k)^2\delta_1\cdots\delta_d$, where the sum is over all cells, does not contribute to the $\delta_j^2\delta_\ell^2$ or δ_j^4 terms, in fact $\sum_k \dot{f}_j(x_k)^2\delta_1\cdots\delta_d = \int(\dot{f}_j)^2dx + O(\delta_1^3+\cdots+\delta_d^3)$, see the Appendix.

(3.8) can be simplified further if f is assumed to go smoothly to zero at infinity so that $\int\frac{\partial}{\partial x_j}\{\dot{f}_j(x)\ddot{f}_{jj}(x)\}dx = 0$ (then $\int \dot{f}_j\,\dddot{f}_{jjj}\,dx = -\int(\ddot{f}_{jj})^2dx$), $\int\frac{\partial}{\partial x_\ell}\{\dot{f}_j(x)\ddot{f}_{j\ell}(x)\}dx = 0$ (then $\int\dot{f}_j\,\dddot{f}_{j\ell\ell}\,dx = -\int(\ddot{f}_{j\ell})^2dx$), $\int\frac{\partial}{\partial x_\ell}\{\ddot{f}_{jj}(x)\dot{f}_\ell(x)\}dx = 0$ (then $\int\ddot{f}_{jj}\ddot{f}_{\ell\ell}dx = \int(\ddot{f}_{j\ell})^2dx$). In this

case

$$\int_{\mathbb{R}^d} \{\text{bias}(x)\}^2 \doteq \sum_{j=1}^{d} \frac{1}{12}\frac{h_j^2}{m_j^2}\int (\dot{f}_j)^2 dx$$
$$+ \sum_{j=1}^{d} \frac{1}{144} h_j^4 \left(1 - \frac{2}{m_j^2} + \frac{3}{5}\frac{1}{m_j^4}\right) \int (\ddot{f}_{jj})^2 dx \tag{3.9}$$
$$+ \sum_{j \neq \ell} \frac{1}{144} h_j^2 h_\ell^2 \left(1 - \frac{1}{m_j^2} - \frac{1}{m_\ell^2} + \frac{1}{2}\frac{1}{m_j^2 m_\ell^2}\right) \int \ddot{f}_{jj} \ddot{f}_{\ell\ell} dx.$$

3.2. Variance of the ASH.

From (3.3), (3.4), (3.5) and using multinomial moment formulae one gets

$$\mathrm{Var}\{\hat f_{\mathrm{ASH}}(x)\} = (h_1\cdots h_d)^{-2}\sum_{i_1,\cdots,i_d} w(i)^2\frac{1}{n}p(i)\{1-p(i)\}$$

$$-(h_1\cdots h_d)^{-2}\sum_{\substack{(i_1,\cdots,i_d)\\ \neq(i'_1,\cdots,i'_d)}} w(i)w(i')\frac{1}{n}p(i)p(i')$$

$$=(nh_1\cdots h_d)^{-1}\sum_i \frac{w(i)^2}{m_1\cdots m_d}\frac{p(i)}{\delta_1\cdots\delta_d}-\frac{1}{n}\left\{\sum_i\frac{w(i)}{m_1\cdots m_d}\frac{p(i)}{\delta_1\cdots\delta_d}\right\}^2$$

$$\doteq (nh_1\cdots h_d)^{-1}\sum_i\frac{w(i)}{m_1\cdots m_d}\prod_{j=1}^d\left(1-\frac{|i_j|}{m_j}\right)\{f(x_0)+\sum_{j=1}^d \dot f_j(x_0)i_j\delta_j\}$$

$$-\frac{1}{n}\Big[\sum_i\frac{w(i)}{m_1\cdots m_d}\{f(x_0)+\sum_{j=1}^d\dot f_j(x_0)i_j\delta_j$$

$$+\frac{1}{2}\sum_{j=1}^d\ddot f_{jj}(x_0)(i_j^2+\frac{1}{12})\delta_j^2+\frac{1}{2}\sum_{j\neq\ell}\ddot f_{j\ell}(x_0)i_ji_\ell\delta_j\delta_\ell\}\Big]^2$$

$$=(nh_1\cdots h_d)^{-1}\prod_{j=1}^d\left(\frac{2}{3}+\frac{1}{3m_j^2}\right)f(x_0)-\frac{1}{n}\left\{f(x_0)+\sum_{j=1}^d\ddot f_{jj}(x_0)\frac{1}{12}(m_j^2-\frac{1}{2})\delta_j^2\right\}^2.$$

This leads to

$$\int_{C_0}\mathrm{Var}\{\hat f_{\mathrm{ASH}}(x)\}dx \doteq (nh_1\cdots h_d)^{-1}(\frac{2}{3})^d\prod_{j=1}^d\left(1+\frac{1}{2m_j^2}\right)f(x_0)\delta_1\cdots\delta_d-\frac{1}{n}f(x_0)^2\delta_1\cdots\delta_d, \tag{3.10}$$

and, combined the the conclusion of the preceding subsection, the IMSE expression reported in Theorem 2 below.

3.3. IMAD for the ASH.

We look for a suitable expression for IMAD $=\int E|\hat f_{\mathrm{ASH}}(x)-E\hat f_{\mathrm{ASH}}(x)|dx$, and again start out trying to evaluate the function over the single cell C_0, on which $\hat f_{\mathrm{ASH}}$ has the constant value $(h_1\cdots h_d)^{-1}\sum_i w(i)Y(i)/n$, cf. (3.2)–(3.4). So

$$\hat f_{\mathrm{ASH}}(x)-E\hat f_{\mathrm{ASH}}(x)=(h_1\cdots h_d)^{-1}\sum_i w(i)\{\frac{1}{n}Y(i)-p(i)\}$$

$$=(h_1\cdots h_d)^{-1}\sum_i w(i)\frac{Y(i)-np(i)}{[np(i)\{1-p(i)\}]^{1/2}}\frac{[p(i)\{1-p(i)\}]^{1/2}}{\sqrt{n}} \tag{3.11}$$

$$=n^{-1/2}(\delta_1\cdots\delta_d)^{-3/2}Z_{0,n},$$

where

$$Z_{0,n} = \sum_i \frac{w(i)}{m_1 \cdots m_d} N(i)[p(i)\{1-p(i)\}]^{1/2}(\delta_1 \cdots \delta_d)^{1/2} \tag{3.12}$$

and

$$N(i) = \{Y(i) - np(i)\}/[np(i)\ \{1-p(i)\}]^{1/2}. \tag{3.13}$$

It follows that

$$\int_{C_0} E|\hat{f}_{\text{ASH}}(x) - E\hat{f}_{\text{ASH}}(x)|dx = (n\delta_1 \cdots \delta_d)^{-1/2} E|Z_{0,n}|. \tag{3.14}$$

Now $Z_{0,n}$ has variance

$$\begin{aligned}
&\sum_i \frac{w(i)^2}{(m_1 \cdots m_d)^2} p(i)\{1-p(i)\}(\delta_1 \cdots \delta_d) - \sum_{i \neq i'} \frac{w(i)w(i')}{(m_1 \cdots m_d)^2} p(i)p(i')(\delta_1 \cdots \delta_d) \\
&\quad = \frac{\delta_1 \cdots \delta_d}{m_1 \cdots m_d} \sum_i \frac{w(i)^2}{m_1 \cdots m_d} p(i) - \left\{ \sum_i \frac{w(i)}{m_1 \cdots m_d} p(i) \right\}^2 \delta_1 \cdots \delta_d \\
&\quad \doteq f(x_0)(\delta_1 \cdots \delta_d)^2 \frac{1}{m_1 \cdots m_d} \sum_i \frac{w(i)^2}{m_1 \cdots m_d} - f(x_0)^2(\delta_1 \cdots \delta_d)^3 \\
&\quad \doteq f(x_0)(\delta_1 \cdots \delta_d)^2 \frac{1}{m_1 \cdots m_d} (\frac{2}{3})^d \prod_{j=1}^d \left(1 + \frac{1}{2m_j^2}\right),
\end{aligned}$$

using (3.5) once more. $np(i) \doteq f(x_0)n\delta_1 \cdots \delta_d \to \infty$ under the standard assumptions. Hence $N(i)$ is asymptotically standard normal, and the $\prod_{j=1}^d (2m_j - 1)$ $N(i)$'s that are involved in $Z_{0,n}$ are asymptotically independent, as the variance computation above also indicates. Hence $Z_{0,n}$ is approximately a zero mean normal random variable, and we should have

$$\begin{aligned}
E|Z_{0,n}| &\doteq (\frac{2}{\pi})^{1/2}(\text{Var } Z_{0,n})^{1/2} \\
&\doteq (\frac{2}{\pi})^{1/2} f(x_0)^{1/2} \delta_1 \cdots \delta_d (m_1 \cdots m_d)^{-1/2} (\frac{2}{3})^{d/2} \prod_{j=1}^d \left(1 + \frac{1}{2m_j^2}\right)^{1/2},
\end{aligned}$$

and accordingly, using (3.14),

$$\int_{C_0} \text{mad}\{\hat{f}_{\text{ASH}}(x)\}dx \doteq (nh_1 \cdots h_d)^{-1/2} (\frac{2}{\pi})^{1/2} (\frac{2}{3})^{d/2} \prod_{j=1}^d \left(1 + \frac{1}{2m_j^2}\right)^{1/2} f(x_0)^{1/2} \delta_1 \cdots \delta_d. \tag{3.15}$$

Summing over all cells

$$\int \operatorname{mad}\{\hat{f}_{\text{ASH}}(x)\}dx \doteq (nh_1\cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2}(\frac{2}{3})^{d/2}\prod_{j=1}^{d}\left(1+\frac{1}{2m_j^2}\right)^{1/2}\int f^{1/2}dx. \tag{3.16}$$

The reasoning above was only suggestive, and several details must be carefully checked in order to establish (3.16) in a precise fashion. These details are dealt with in the proof of Theorem 2 in the Appendix.

<u>3.4. IAB for the ASH.</u>

We already have the expression (3.7) for $\text{bias}(x) = E\hat{f}_{\text{ASH}}(x) - f(x)$. If $m_j \geq 2$ then $(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 > \frac{1}{2}(x_j - x_{0,j})^2$ when $x \in C_0$. Hence

$$\begin{aligned}
\int_{C_0} |\text{bias}(x)|dx \leq & \int_{C_0} |\sum_{j=1}^{d} \dot{f}_j(x_0)(x_j - x_{0,j})|dx \\
& + \int_{C_0} |\sum_{j=1}^{d} \ddot{f}_{jj}(x_0)\{(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 - \frac{1}{2}(x_j - x_{0,j})^2\}|dx \\
& + \frac{1}{2}\int_{C_0} |\sum_{j\neq \ell} \ddot{f}_{j\ell}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})|dx \\
& + \frac{1}{6}\int_{C_0} |\sum_{j,\ell,s} \dddot{f}_{j\ell s}(x_0)(x_j - x_{0,j})(x_\ell - x_{0,\ell})(x_s - x_{0,s})|dx \\
& + O\left(\sum_{i=1}^{d}\delta_j^4\right)\delta_1\cdots\delta_d \leq \sum_{j=1}^{d}|\dot{f}_j(x_0)|\frac{1}{4}\delta_j\ \delta_1\cdots\delta_d \\
& + \sum_{j=1}^{d}|\ddot{f}_{jj}(x_0)|\{(m_j^2 - \frac{1}{2})\frac{1}{12}\delta_j^2 - \frac{1}{24}\delta_j^2\}\delta_1\cdots\delta_d \\
& + \frac{1}{2}\sum_{j\neq \ell}|\ddot{f}_{j\ell}(x_0)|\frac{1}{4}\delta_j\frac{1}{4}\delta_\ell\ \delta_1\cdots\delta_d \\
& + O(\delta_1^3 + \cdots + \delta_d^3)\sum_{j,\ell,s}|\dddot{f}_{j\ell s}(x_0)|\delta_1\cdots\delta_d.
\end{aligned}$$

Accordingly

$$\begin{aligned}
\text{IAB } & \leq \sum_{j=1}^{d}\frac{1}{4}\delta_j\int|\dot{f}_j|dx + \sum_{j=1}^{d}\frac{1}{12}\delta_j^2(m_j^2 - 1)\int|\ddot{f}_{jj}|dx + \sum_{j\neq \ell}\frac{1}{32}\delta_j\delta_\ell\int|\ddot{f}_{j\ell}|dx \\
& = \sum_{j=1}^{d}\{\frac{1}{4}\frac{h_j}{m_j}\int|\dot{f}_j|dx + \frac{1}{12}h_j^2\left(1 - \frac{1}{m_j^2}\right)\int|\ddot{f}_{jj}|dx\} + \sum_{j\neq \ell}\frac{1}{32}\frac{h_j}{m_j}\frac{h_\ell}{m_\ell}\int|\ddot{f}_{j\ell}|dx,
\end{aligned}$$

ignoring higher order terms.

Theorem 2. Let f have continuous partial derivatives to the fourth order, and assume that all the functions $|\dot{f}_j|$, $|\ddot{f}_{j\ell}|$, $|\dddot{f}_{j\ell s}|$, $|\ddddot{f}_{j\ell st}|$, and their squares are integrable. Assume also that f goes to zero at infinity smoothly enough to ensure $\int (\ddot{f}_{j\ell})^2 dx = \int \ddot{f}_{jj}\ddot{f}_{\ell\ell} dx = -\int \dot{f}_j \dddot{f}_{j\ell\ell}\, dx$. Then for Scott's average shifted histogram defined in (3.2) and (3.3), as $h_1 = m_1\delta_1 \to 0, \cdots, h_d = m_d\delta_d \to 0$, $n\delta_1 \cdots \delta_d \to \infty$,

$$\begin{aligned}
\text{IMSE} &= (nh_1\cdots h_d)^{-1}(\tfrac{2}{3})^d \prod_{j=1}^{d}\left(1+\frac{1}{2m_j^2}\right) - \frac{1}{n}\int f^2 dx \\
&+ \sum_{j=1}^{d} \frac{1}{12}\frac{h_j^2}{m_j^2}\int (\dot{f}_j)^2 dx + \sum_{j=1}^{d}\frac{1}{144}h_j^4\left(1-\frac{2}{m_j^2}+\frac{3}{5}\frac{1}{m_j^4}\right)\int (\ddot{f}_{jj})^2 dx \\
&+ \sum_{j\neq\ell}\frac{1}{144}h_j^2h_\ell^2\left(1-\frac{1}{m_j^2}-\frac{1}{m_\ell^2}+\frac{1}{2}\frac{1}{m_j^2m_\ell^2}\right)\int \ddot{f}_{jj}\ddot{f}_{\ell\ell}dx \\
&+ O\left(\sum_{j=1}^{d}\frac{h_j^5}{m_j} + \sum_{j=1}^{d}\frac{h_j^2}{nh_1\ldots h_d}\right).
\end{aligned}$$

Furthermore, over each bounded region where $\int |\ddot{f}_{j\ell}|/f^{1/2}dx$ is finite,

$$\begin{aligned}
\text{IMAD} + \text{IAB} &\leq (nh_1\cdots h_d)^{-1/2}(\tfrac{2}{\pi})^{1/2}(\tfrac{2}{3})^{d/2}\prod_{j=1}^{d}\left(1+\frac{1}{2m_j^2}\right)^{1/2}\int f^{1/2}dx \\
&+ \sum_{j=1}^{d}\frac{1}{4}\frac{h_j}{m_j}\int |\dot{f}_j| dx + \sum_{j=1}^{d}\frac{1}{12}h_j^2\left(1-\frac{1}{m_j^2}\right)\int |\ddot{f}_{jj}| dx \\
&+ \sum_{j\neq\ell}\frac{1}{32}\frac{h_j}{m_j}\frac{h_\ell}{m_\ell}\int |\ddot{f}_{j\ell}| dx + O\left(\sum_{j=1}^{d}\frac{h_j^3}{m_j} + \sum_{j=1}^{d}\frac{h_j^2}{nh_1\cdots h_d} + \frac{m_1^2\cdots m_d^2}{nh_1\cdots h_d}\right).
\end{aligned}$$

4. Generalized Frequency Polygons of Average Shifted Histograms

The average shifted histogram is

$$\hat{f}_{\text{ASH}}(x) = (h_1\cdots h_d)^{-1}\sum_{i_1,\cdots,i_d} w(i_1,\cdots,i_d)Y(i_1,\cdots,i_d)/n$$

for each $x \in C_0 = [x_0 - \frac{1}{2}\delta,\ x_0 + \frac{1}{2}\delta]$, where $Y(i_1,\cdots,i_d) = Y(i)$ is the number of observations falling in $C_0(i_1,\cdots,i_d) = C_0(i) = (x_0 + (i-\frac{1}{2})\delta,\ x_0 + (i+\frac{1}{2})\delta]$. In order to construct the generalized frequency polygon of $\hat{f}_{\text{ASH}}$ we should interpolate between 2^d neighbour values of $\hat{f}_{\text{ASH}}$. Consider therefore the 2^d ASH-cells $C_0(j_1,\cdots,j_d) = (x_0 + (j-\frac{1}{2})\delta,\ x_0 + (j+\frac{1}{2})\delta]$,

$j_1, \cdots, j_d \in \{0, 1\}$, and shift attention to the inner cell

$$C_0^* = (x_0,\ x_0 + \delta] = (a - \frac{1}{2}\delta,\ a + \frac{1}{2}\delta], \tag{4.1}$$

where $a = x_0 + \frac{1}{2}\delta$ is the centre point of this new GFP-ASH cell and at the same time in the middle of 2^d old ASH-cells.

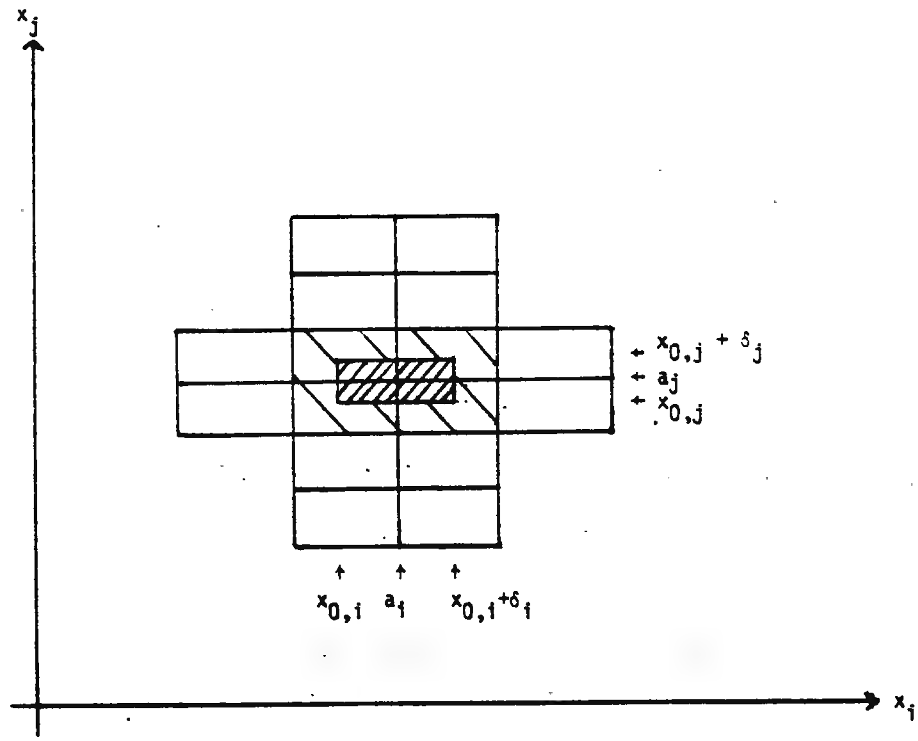


Figure 2. The GFP-ASH cell lies within 2^d ASH cells. $\hat{f}_{\text{ASH}}(x)$ is defined, for each of these cells, as a weighted average over $(2m_1 - 1) \cdots (2m_d - 1)$ ASH cells. This is illustrated with $m_i = 2$ and $m_j = 3$ above.

Define, then, for x in C_0^*,

$$\begin{aligned}
\hat{f}_{\text{GFP-ASH}}(x) &= \sum_{j_1, \cdots, j_d} c_{j_1, \cdots, j_d}(x) \hat{f}_{\text{ASH}}(x_{0,1} + j_1\delta_1, \cdots, x_{0,d} + j_d\delta_d) \\
&= (h_1 \cdots h_d)^{-1} \sum_{j_1, \cdots, j_d} \sum_{i_1 = 1 - m_1}^{m_1 - 1} \cdots \sum_{i_d = 1 - m_d}^{m_d - 1} w(i_1, \cdots, i_d) Y(j_1 + i_1, \cdots, j_d + i_d)/n \\
&= (h_1 \cdots h_d)^{-1} \sum_{\ell_1 = 1 - m_1}^{m_1} \cdots \sum_{\ell_d = 1 - m_d}^{m_d} T(\ell_1, \cdots, \ell_d) Y(\ell_1, \cdots, \ell_d)/n,
\end{aligned} \tag{4.2}$$

where $w(i) = \prod_{v=1}^{d} \left(1 - \frac{|i_v|}{m_v}\right)$ as in Section 3, where

$$c_j(x) = (1-u_1)^{1-j_1} u_1^{j_1} \cdots (1-u_d)^{1-j_d} u_d^{j_d} \tag{4.3}$$

as in Section 2, but now with

$$u_v = u_v(x) = (x_v - x_{0,v})/\delta_v = \frac{1}{2} + (x_v - a_v)/\delta_v, \qquad v = 1, \cdots, d, \tag{4.4}$$

and where finally

$$\begin{aligned} T(\ell) &= T(\ell_1, \cdots, \ell_d) \\ &= \sum_{\substack{i_1,\cdots,i_d \\ i_v + j_v = \ell_v,\, v=1,\cdots,d}} \sum_{j_1,\cdots,j_d} w(i_1, \cdots, i_d) c_{j_1,\cdots,j_d}(x) \\ &= T_1(\ell_1) \cdots T_d(\ell_d), \end{aligned} \tag{4.5}$$

with

$$\begin{aligned} T_v(\ell_v) &= \sum_{\substack{1-m_v \le i_v \le m_v - 1 \\ 0 \le j_v \le 1 \\ i_v + j_v = \ell_v}} \left(1 - \frac{|i_v|}{m_v}\right) (1-u_v)^{1-j_v} u_v^{j_v} \\ &= (1-u_v)\left(1 - \frac{|\ell_v|}{m_v}\right) + u_v \left(1 - \frac{|\ell_v - 1|}{m_v}\right), \quad \ell_v = 1 - m_v, \cdots, m_v. \end{aligned} \tag{4.6}$$

4.1. Bias of the GFP-ASH.

The expectation and variance of $\hat{f}_{\text{GFP-ASH}}$ depend upon the cell probabilities $p(\ell) = \int_{C_0(\ell)} f \, dx$. These were studied in (3.5), but it is now more advantageous to Taylor expand f

around $x = a$:

$$
\begin{aligned}
p(\ell) &= \int_{(a+(\ell-1)\delta, a+\ell\delta]} \Big\{ f(a) + \sum_{i=1}^{d} \dot{f}_i(a)(x_i - a_i) + \frac{1}{2}\sum_{i,j} \ddot{f}_{ij}(a)(x_i - a_i)(x_j - a_j) \\
&\quad + \frac{1}{6}\sum_{i,j,s} \dddot{f}_{ijs}\,(a)(x_i - a_i)(x_i - a_j)(x_s - a_s) + \cdots \Big\} dx \\
&\doteq f(a)\delta_1 \cdots \delta_d + \sum_{i=1}^{d} \dot{f}_i(a) \int_{((\ell-1)\delta, \ell\delta]} x_i dx \\
&\quad + \frac{1}{2}\sum_{i=1}^{d} \ddot{f}_{ii}(a) \int_{((\ell-1)\delta, \ell\delta]} x_i^2 dx + \frac{1}{2}\sum_{i \neq j} \ddot{f}_{ij}(a) \int_{((\ell-1)\delta, \ell\delta]} x_i x_j \, dx \qquad (4.7) \\
&= f(a)\delta_1 \cdots \delta_d + \sum_{i=1}^{d} \dot{f}_i(a)(\ell_i - \frac{1}{2})\delta_i \;\; (\delta_1 \cdots \delta_d) \\
&\quad + \sum_{i=1}^{d} \frac{1}{2}\ddot{f}_{ii}(a)(\ell_i^2 - \ell_i + \frac{1}{3})\delta_i^2 \;\; (\delta_1 \cdots \delta_d) \\
&\quad + \sum_{i \neq j} \frac{1}{2}\ddot{f}_{ij}(a)(\ell_i - \frac{1}{2})\delta_i(\ell_j - \frac{1}{2})\delta_j \;\; (\delta_1 \cdots \delta_d).
\end{aligned}
$$

Some exercises in algebra yield $\sum_{\ell_i=1-m_i}^{m_i} T_i(\ell_i) = m_i$, $\sum_{\ell_i=1-m_i}^{m_i} (\ell_i - \frac{1}{2})T_i(\ell_i) = (u_i - \frac{1}{2})m_i$, and $\sum_{\ell_i=1-m_i}^{m_i} (\ell_i^2 - \ell_i + \frac{1}{3})T_i(\ell_i) = \frac{1}{6}(m_i^2 + 1)m_i$. All this leads to

$$
\begin{aligned}
E\hat{f}_{\text{GFP-ASH}}(x) &= (h_1 \cdots h_d)^{-1} \sum_{\ell_1, \cdots, \ell_d} T(\ell_1, \cdots, \ell_d) p(\ell_1, \cdots, \ell_d) \\
&\doteq \frac{\delta_1 \cdots \delta_d}{h_1 \cdots h_d} \Big[f(a) m_1 \cdots m_d + \sum_{i=1}^{d} \dot{f}_i(a)(u_i - \frac{1}{2})\delta_i \; m_1 \cdots m_d \\
&\quad + \sum_{i=1}^{d} \frac{1}{2}\ddot{f}_{ii}(a)\frac{1}{6}(m_i^2 + 1)\delta_i^2 \; m_1 \cdots m_d \\
&\quad + \sum_{i \neq j} \frac{1}{2}\ddot{f}_{ij}(a)(u_i - \frac{1}{2})(u_j - \frac{1}{2})\delta_i\delta_j \; m_1 \cdots m_d \Big] \\
&= f(a) + \sum_{i=1}^{d} \dot{f}_i(a)(x_i - a_i) + \sum_{i=1}^{d} \frac{1}{2}\ddot{f}_{ii}(a)\frac{1}{6}(m_i^2 + 1)\delta_i^2 \\
&\quad + \frac{1}{2}\sum_{i \neq j} \ddot{f}_{ij}(a)(x_i - a_i)(x_j - a_j),
\end{aligned}
$$

using (4.4) and remembering $\delta_j = h_j/m_j$, $j = 1, \cdots, d$. Subtracting a Taylor approximation

for $f(x)$ we have arrived at

$$\begin{aligned}\text{bias}(x) &= E\hat{f}_{\text{GFP-ASH}}(x) - f(x) \\ &\doteq \sum_{i=1}^{d} \ddot{f}_{ii}(a)\{\frac{1}{12}(m_i^2+1)\delta_i^2 - \frac{1}{2}(x_i - a_i)^2\}.\end{aligned} \tag{4.8}$$

We could actually have reached (4.8) more directly, perhaps recycling the efforts of Section 3 better, cf. (3.5) and (3.6), but the representation (4.2) is in any case needed in Section 4.2 below.

From (4.8) one gets

$$\begin{aligned}\int_{(a-\frac{1}{2}\delta,\, a+\frac{1}{2}\delta]} \{\text{bias}(x)\}^2 dx &\doteq \sum_{i=1}^{d} \ddot{f}_{ii}(a)^2 \int_{(-\frac{1}{2}\delta,\, \frac{1}{2}\delta]} \{\frac{1}{12}(m_i^2+1)\delta_i^2 - \frac{1}{2}x_i^2\}^2 dx \\ &\quad + \sum_{i\neq j} \ddot{f}_{ij}(a)\ddot{f}_{jj}(a) \int_{(-\frac{1}{2}\delta,\, \frac{1}{2}\delta]} \{\frac{1}{12}(m_i^2+1)\delta_i^2 - \frac{1}{2}x_i^2\}\{\frac{1}{12}(m_j^2+1)\delta_j^2 - \frac{1}{2}x_j^2\} dx \\ &= \sum_{i=1}^{d} \ddot{f}_{ii}(a)^2 \frac{1}{144}\delta_i^4 (m_i^4 + m_i^2 + \frac{9}{20})(\delta_1 \cdots \delta_d) \\ &\quad + \sum_{i\neq j} \ddot{f}_{ii}(a)\ddot{f}_{jj}(a)\frac{1}{144}\delta_i^2\delta_j^2 (m_i^2 + \frac{1}{2})(m_j^2+\frac{1}{2})(\delta_1\cdots\delta_d).\end{aligned}$$

Summing over all cells and approximating integrals with Riemann sums as usual, and using $\delta_j = h_j/m_j$, we reach

$$\begin{aligned}\int_{\mathbb{R}^d} \{\text{bias}(x)\}^2 dx &\doteq \sum_{i=1}^{d} \frac{1}{144} h_i^4 \left(1 + \frac{1}{m_i^2} + \frac{9/20}{m_i^4}\right) \int (\ddot{f}_{ii})^2 dx \\ &\quad + \sum_{i\neq j} \frac{1}{144} h_i^2 h_j^2 \left(1 + \frac{1}{2m_i^2}\right)\left(1 + \frac{1}{2m_j^2}\right) \int \ddot{f}_{ii}\ddot{f}_{jj} dx \\ &= \frac{1}{144} \int \left\{\sum_{i=1}^{d} h_i^2 \left(1 + \frac{1}{2m_i^2}\right) \ddot{f}_{ii}\right\}^2 dx + \sum_{i=1}^{d} \frac{1}{720}\frac{h_i^4}{m_i^4} \int (\ddot{f}_{ii})^2 dx.\end{aligned} \tag{4.9}$$

4.2. Variance of the GFP-ASH.

$\hat{f}_{\text{GFP-ASH}}(x)$ is a linear combination of $\hat{f}_{\text{ASH}}(x_0 + j\delta)$ values, and its variance involves $\text{cov}\{\hat{f}_{\text{ASH}}(x_0 + j\delta),\, \hat{f}_{\text{ASH}}(x_0 + j'\delta)\}$, $j, j' \in \{0,1\}^d$. It is most convenient to use the repre-

sentation (4.2). We get, for $x \in (a - \frac{1}{2}\delta,\ a + \frac{1}{2}\delta]$,

$$
\begin{aligned}
\mathrm{Var}\{\hat{f}_{\text{GFP-ASH}}(x)\} &= n^{-1}(h_1\cdots h_d)^{-2}\sum_{\ell_1,\cdots,\ell_d} T(\ell)^2 p(\ell)\{1-p(\ell)\} \\
&\quad - n^{-1}(h_1\cdots h_d)^{-2}\sum_{\ell\neq\ell'} T(\ell)T(\ell')p(\ell)p(\ell') \qquad (4.10)\\
&= n^{-1}(h_1\cdots h_d)^{-1}\sum_{\ell}\frac{T(\ell)^2}{m_1\cdots m_d}\frac{p(\ell)}{\delta_1\cdots\delta_d} - n^{-1}\{(h_1\cdots h_d)^{-1}\sum_{\ell}T(\ell)p(\ell)\}^2.
\end{aligned}
$$

One may show that

$$
\begin{aligned}
\sum_{\ell_i=1-m_i}^{m_i} T_i(\ell_i)^2 &= (1-u_i)^2\sum_{\ell=1-m_i}^{m_i}\left(1-\frac{|\ell|}{m_i}\right)^2 + u_i^2\sum_{\ell=1-m_i}^{m_i}\left(1-\frac{|\ell-1|}{m_i}\right) \\
&\quad + 2u_i(1-u_i)\sum_{\ell=1-m_i}^{m_i}\left(1-\frac{|\ell|}{m_i}\right)\left(1-\frac{|\ell-1|}{m_i}\right) \\
&= \{2m_i^2+1-6u_i(1-u_i)\}/3m_i.
\end{aligned}
$$

Hence, since $p(\ell)/(\delta_1\cdots\delta_d) \doteq f(a)$ by (4.7),

$$
\begin{aligned}
\int_{(a-\frac{1}{2}\delta,\ a+\frac{1}{2}\delta]}\sum_{\ell}\frac{T(\ell)^2}{m_1\cdots m_d}\frac{p(\ell)}{\delta_1\cdots\delta_d}dx &\doteq \int_{[a-\frac{1}{2}\delta,\ a+\frac{1}{2}\delta]}\prod_{i=1}^{d}\frac{2m_i^2+1-6u_i(1-u_i)}{3m_i^2}dx\ f(a) \\
&= \prod_{i=1}^{d}\{(2m_i^2+1-1)/3m_i^2\}\delta_i\ f(a) \\
&= (\frac{2}{3})^d f(a)\ \delta_1\cdots\delta_d.
\end{aligned}
\qquad (4.11)
$$

Also, $\sum_{\ell} T(\ell)p(\ell) \doteq m_1\cdots m_d\ f(a)\ \delta_1\cdots\delta_d$, so that $\{(h_1\cdots h_d)^{-1}\sum_{\ell}T(\ell)p(\ell)\}^2 \doteq f(a)^2$. It follows that

$$
\int_{(a-\frac{1}{2}\delta,\ a+\frac{1}{2}\delta]}\mathrm{Var}\{\hat{f}_{\text{GFP-ASH}}(x)\}dx \doteq (\frac{2}{3})^d(nh_1\cdots h_d)^{-1} - \frac{1}{n}\int f^2 dx. \qquad (4.12)
$$

4.3. IMAD of the GFP-ASH.

From (4.2)

$$
\begin{aligned}
\hat{f}_{\text{GFP-ASH}}(x) - E\hat{f}_{\text{GFP-ASH}}(x) &= (h_1\cdots h_d)^{-1}\sum_{\ell}T(\ell)\{\frac{1}{n}Y(\ell)-p(\ell)\} \\
&= (nh_1\cdots h_d)^{-1/2}\sum_{\ell}T(\ell)\frac{Y(\ell)-np(\ell)}{[np(\ell)\{1-p(\ell)\}]^{1/2}}\frac{[p(\ell)\{1-p(\ell)\}]^{1/2}}{(h_1\cdots h_d)^{1/2}} \\
&= (nh_1\cdots h_d)^{-1/2}Z(x),
\end{aligned}
$$

say, $x \in (a - \frac{1}{2}\delta,\ a + \frac{1}{2}\delta]$, where, reasoning as in 2.4 and 3.3, $Z(x)$ is approximately normal with zero mean and variance

$$\sigma^2(x) = \sum_{\ell} T(\ell)^2 \frac{p(\ell)}{h_1 \cdots h_d} = \sum_{\ell} \frac{T(\ell)^2}{m_1 \cdots m_d} \frac{p(\ell)}{\delta_1 \cdots \delta_d}$$
$$\doteq f(a) \prod_{i=1}^{d} \{2m_i^2 + 1 - 6u_i(1-u_i)\}/(3m_i^2).$$

This suggests $E|Z(x)| \doteq (\frac{2}{\pi})^{1/2}\sigma(x)$ and

$$\int_{(a-\frac{1}{2}\delta,\ a+\frac{1}{2}\delta]} E|\hat{f}_{\text{GFP-ASH}}(x) - E\hat{f}_{\text{GFP-ASH}}(x)|dx$$
$$\doteq (nh_1 \cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2} f(a)^{1/2} \prod_{i=1}^{d} \int_{a_i-\frac{1}{2}\delta_i}^{a_i+\frac{1}{2}\delta_i} [\{2m_i^2 + 1 - 6u_i(1-u_i)\}/(3m_i^2)]^{1/2} dx_i \quad (4.13)$$
$$= (nh_1 \cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2} f(a)^{1/2} (\delta_1 \cdots \delta_d)\ J(m_1) \cdots J(m_d),$$

where

$$J(m_i) = \int_0^1 \{2m_i^2 + 1 - 6u(1-u)\}^{1/2} du/\sqrt{3}m_i$$
$$= (\frac{1}{6} + \frac{1}{12m_i^2})^{1/2} + \frac{4m_i^2 - 1}{6\sqrt{2}m_i} \log \frac{3^{1/2} + (4m_i^2 + 2)^{1/2}}{(4m_i^2 - 1)^{1/2}}. \quad (4.14)$$

($J(m_i)$ is amazingly close to its limit value $(\frac{2}{3})^{1/2}$ already for $m_i = 3$.) Accordingly

$$\text{IMAD} \doteq (nh_1 \cdots h_d)^{-1}(\frac{2}{\pi})^{1/2} \int f^{1/2} dx\ J(m_1) \cdots J(m_d). \quad (4.15)$$

4.4. IAB of the GFP-ASH.

From (4.8) it is clear that

$$\int_{(a-\frac{1}{2}\delta,\, a+\frac{1}{2}\delta]} |\text{bias}(x)|dx \le \sum_{i=1}^{d} |\ddot{f}_{ii}(a)| \int_{(-\frac{1}{2}\delta,\, \frac{1}{2}\delta]} \{\frac{1}{12}(m_i^2+1)\delta_i^2 - \frac{1}{2}x_i^2\}dx$$
$$= \sum_{i=1}^{d} |\ddot{f}_{ii}(a)|\, \delta_1 \cdots \delta_d\, \frac{1}{12}\delta_i^2(m_i^2+\frac{1}{2}),$$

and in its turn

$$\text{IAB} \le \sum_{i=1}^{d} \frac{1}{12}h_i^2\left(1+\frac{1}{2m_i^2}\right)\int |\ddot{f}_{ii}|dx. \tag{4.16}$$

Some of the details that remain in order to actually prove the following theorem are available in the Appendix.

Theorem 3. For the generalized frequency polygon of the average shifted histogram:

$$\begin{aligned}
\text{IMSE} &= (\frac{2}{3})^d(nh_1\cdots h_d)^{-1} - \frac{1}{n}\int f^2dx \\
&+ \sum_{i=1}^{d}\frac{1}{144}h_i^4\left(1+\frac{1}{m_i^2}+\frac{9/20}{m_i^4}\right)\int(\ddot{f}_{ii})^2dx \\
&+ \sum_{i<j}\frac{1}{72}h_i^2h_j^2\left(1+\frac{1}{2m_i^2}\right)\left(1+\frac{1}{2m_j^2}\right)\int \ddot{f}_{ii}\ddot{f}_{jj}dx + O\left(\sum_{i=1}^{d}\frac{h_i^5}{m_i}+\sum_{i=1}^{d}\frac{h_i^2}{nh_1\cdots h_d}\right),
\end{aligned}$$

$$\begin{aligned}
\text{IMAD} + \text{IAB} &\le (nh_1\cdots h_d)^{-1/2}(\frac{2}{\pi})^{1/2}\int f^{1/2}dx\, J(m_1)\cdots J(m_d) \\
&+ \sum_{i=1}^{d}\frac{1}{12}h_i^2\left(1+\frac{1}{2m_i^2}\right)\int|\ddot{f}_{ii}|dx \\
&+ O\left(\sum_{i=1}^{d}\frac{h_i^3}{m_i}+\sum_{i=1}^{d}\frac{h_i^2}{(nh_1\cdots h_d)^{1/2}}+\frac{1}{nh_1\cdots h_d}\right).
\end{aligned}$$

It is assumed that f has third order continuous derivatives. The IMSE expression holds when f^2, $(\dot{f}_i)^2$, $(\ddot{f}_{ij})^2$, $(\dddot{f}_{ijk})^2$ have finite integrals. The IMAD + IAB expression holds over each bounded region where f, $|\dot{f}_i|$, $|\ddot{f}_{ij}|$, $|\dddot{f}_{ijk}|$, $|\ddot{f}_{ij}|/f^{1/2}$, $|\dot{f}_i\dot{f}_j|/f^{3/2}$ all have finite integrals.

We remark that the IMSE expression obtained here is better than the one obtained (for $d=2$) in Scott (1985b, Theorem 4), for his version of bivariate frequency polygons of average shifted histograms.

5. Discussion.

We have obtained results for IMSE and for IMAD + IAB for natural generalizations of histograms. Some consequences of these results will be briefly discussed in this section and comparisons with kernel type density estimators will be made.

The papers of Scott (1985a, 1985b) have provided the inspiration for the present paper. We have been able to improve and generalize his results somewhat, but the basic statistical and practical issues remain the same, and Scott's discussion of these points are valid also for this paper's density estimators, with few exceptions and minor modifications. The reader is therefore referred to the above mentioned articles for fuller discussion.

5.1. Comparison with kernel density estimators.

The kernel density estimators are the most usual alternatives to histograms. They are of the form

$$f^*(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_1 - X_{i,1}}{h_1}, \cdots, \frac{x_d - X_{i,d}}{h_d}\right) / (h_1 \cdots h_d), \tag{5.1}$$

with the kernel K a function on $I\!R^d$. Usual requirements are that K is nonnegative and integrates to one, and that it is symmetric; $K(u_1, \cdots, u_d) = K(\epsilon_1 u_1, \cdots, \epsilon_d u_d)$ for all $\epsilon_1, \cdots, \epsilon_d \in \{-1, 1\}$. It is also customary to have $h_1 = \cdots = h_d$, and to employ product kernels, i.e. $K(u_1, \cdots, u_d) = K_0(u_1) \cdots K_0(u_d)$ for some univariate kernel K_0.

It is interesting to compare the results of Theorems 1, 2, and 3 with corresponding expressions for a kernel density estimator. If K is nonnegative with integral one, and $\int u_i K(u)du = 0$, $\tau_i^2 = \int u_i^2 K(u)du$, $\int u_i u_j K(u)du = 0$ for $i \neq j$, and $\int K(u)^2 du$ is finite, then one can show that

$$\begin{aligned} \text{IMSE} &\doteq \int K^2 du\,(nh_1 \cdots h_d)^{-1} - \frac{1}{n}\int f^2 dx \\ &\quad + \sum_{i=1}^{d} \frac{1}{4} h_i^4 \tau_i^4 \int (\ddot{f}_{ii})^2 + \sum_{i \neq j} \frac{1}{4} h_i^2 h_j^2 \int \ddot{f}_{ii} \ddot{f}_{jj} dx, \end{aligned} \tag{5.2}$$

$$\text{IMAD + IAB} \leq \left(\frac{2}{\pi}\right)^{1/2} \left(\int K^2 du\right)^{1/2} \int f^{1/2} dx\,(nh_1 \cdots h_d)^{-1/2} + \sum_{i=1}^{d} \frac{1}{2} h_i^2 \tau_i^2 \int |\ddot{f}_{ii}| dx. \tag{5.3}$$

(5.2) is a natural generalization of the classical univariate result of Parzen (1962); see also Epanechnikov (1969).

Let us also write down the results for the ordinary histogram $\hat{f}_0$ of (1.1):

$$\text{IMSE} \doteq (nh_1 \cdots h_d)^{-1} - \frac{1}{n}\int f^2 dx + \sum_{i=1}^{d} \frac{1}{12} h_i^2 \int (\dot{f}_i)^2 dx, \tag{5.4}$$

$$\text{IMAD} + \text{IAB} \leq (\frac{2}{\pi})^{1/2} \int f^{1/2} dx \, (nh_1 \cdots h_d)^{-1/2} + \sum_{i=1}^{d} \frac{1}{4} h_i \int |\dot{f}_i| dx. \tag{5.5}$$

These can be obtained by the methods of Section 2.

The advantage of passing from $\hat{f}_0$ to the more sophisticated kernel estimator is that the bias is of a smaller order. The rate at which IMSE for the histogram goes to zero when the best values for $h_1, \cdots, h_d$ are used is $n^{-2/(d+2)}$, whereas IMSE of (5.2) can attain the rate $n^{-4/(d+4)}$. The same phenomenon is illustrated using the IMAD + IAB criterion. If $h_i = a_i n^{-\alpha}$, $i = 1, \cdots, d$, then the best choice for α in the histogram case (5.5) can be seen to be $1/(d+2)$, and IMAD + IAB has rate $n^{-1/(d+2)}$. For the kernel estimator case (5.3) $\alpha = 1/(d+4)$ is best, giving IMAD + IAB a rate of $n^{-2/(d+4)}$.

It is clear from these considerations and from Theorems 1, 2, and 3 that both the GFP and the GFP-ASH achieve the same favourable rate as the kernel estimator, i.e. $n^{-4/(d+4)}$ for the expected L_2 distance and $n^{-2/(d+4)}$ for the upper bound for the L_1 distance. They therefore offer substantial (asymptotic) improvement over the ordinary histogram. The ASH does not quite achieve the same kernel estimator rates, but for the finite n statisticians are faced with the constants accompanying $n^{-2/(d+2)}$ and $n^{-4/(d+4)}$ determine everything, and it is evident from Theorem 2 that the ASH produces IMSE and IMAD + IAB that match those of the kernel estimator, i.e. (5.2) and (5.3), even for moderate values of m_i. This is no coincidence; Scott (1985b) observes that the ASH of (3.3) is close to the kernel estimator (5.1) with $K(u) = \prod_{i=1}^{d}(1 - |u_i|) I\{|u_i| \leq 1\}$, the product triangle kernel. Indeed (5.2) and (5.3) result if we let the m_i's grow to infinity in the expressions of Theorem 2.

$\hat{f}_{\text{ASH}}$ of (3.3) can be seen as a computationally convenient approximation to f^* of (5.1) with K the product triangle kernel. This points to the possibility of approximating other kernel density estimators in the same manner, using a different weighting scheme than (3.4). Such an approximation works directly with the binned data, and the computational burden is almost independent of the sample size n of the raw data. In an example Scott (1985b) reports

on this meant reducing CPU time from hours to minutes.

Scott (1985b) notes that even small to moderate values of the m_i's are effective in eliminating the portion of the bias that stems from binning the data.

Thus $\hat{f}_{\mathrm{ASH}}$, and relatives, are convenient approximations to kernel estimators, and are as such similar in spirit to the discrete Fourier inversion method introduced in Silverman (1982) (see also Jones and Lotwick (1984)). This technique works for a Gaussian kernel. It is unclear how practical that method is for high-dimensional data.

The GFP-ASH is an interpolated version of a kernel estimator approximation for binned data; see Scott (1985b, Section 6) for further comments.

Finally, it should be pointed out in this subsection that $\hat{f}_0$, $\hat{f}_{\mathrm{GFP}}$, $\hat{f}_{\mathrm{ASH}}$, and $\hat{f}_{\mathrm{GFP\text{-}ASH}}$ in fact all are kernel estimators, but with complicated kernels, being only piecewise continuous, and not all of them symmetric; see Walter and Blum (1979) and Scott (1985b, Section 3). Another way of obtaining Theorems 1, 2, and 3 would conceivably be to determine these underlying kernels explicitly, then prove precise versions of (5.2), (5.3), but for piecewise continuous and nonsymmetric kernels, and then evaluate the appropriate terms.

5.2. Choice of smoothing parameters.

It is natural to choose smoothing parameters so as to minimize the leading terms of either the IMSE or the IMAD + IAB, see for example Freedman and Diaconis (1981) or Scott (1985a).

Consider, for example, the generalized frequency polygon of Section 2, and let us for the moment adapt the L_1 view that led to IMAD + IAB as a natural criterion. The leading terms are of the form

$$A_0(nh_1\cdots h_d)^{-1/2}+\sum_{i=1}^{d}A_ih_i^2,$$

for constants $A_0, A_1, \cdots, A_d$ determined by f. Setting partial derivatives equal to zero one finds that the best choice for $h_1, \cdots, h_d$ is

$$h_i = h_i^* = (\frac{1}{2})^{4/(d+4)}A_0^{2/(d+4)}(A_1\cdots A_d)^{1/2(d+4)}A_i^{-1/2}n^{-1/(d+4)}. \tag{5.6}$$

With this choice,

$$\begin{aligned} \text{IMAD} + \text{IAB} \le \{ & 2^{2d/(d+4)} \\ & + d(\frac{1}{2})^{8/(d+4)}\} A_0^{4/(d+4)} (A_1 \cdots A_d)^{1/(d+4)} n^{-2/(d+4)} + O(n^{-3/(d+4)}). \end{aligned} \tag{5.7}$$

Since A_0 is proportional to $\int f^{1/2} dx$ and A_i to $\int |\ddot{f}_{ii}| dx$, it emerges that

$$\Delta(f) = \left\{ (\int f^{1/2} dx)^4 \int |\ddot{f}_{11}| dx \cdots \int |\ddot{f}_{dd}| dx \right\}^{1/(d+4)} \tag{5.8}$$

is a natural measure of how difficult a particular f is to estimate using a generalized frequency polygon.

Exactly the same reasoning is valid for the GFP-ASH and for the kernel estimator f^*. In both cases h_i should again be taken proportional to $n^{-1/(d+4)}$, and again $\Delta(f)$ of (5.8) appears as a natural measure of the difficulty with which f can be estimated.

Of course $A_0, A_1, \cdots, A_d$ will be unknown, and a natural way to proceed is to estimate the needed quantities $\int f^{1/2} dx$, $\int |\ddot{f}_{11}| dx, \cdots, \int |\ddot{f}_{dd}| dx$ based on the observed data, and plug in estimates in (5.6). The estimation can be performed nonparametrically, using perhaps a separate kernel estimate or spline type estimate of f, and perhaps with separately determined smoothing parameters, for this purpose. Another possibility is to fit perhaps a rough parametric model to the data, and estimate $\int f^{1/2} dx$ etc. using parametric techniques.

To get a possible benchmark for the choice of $h_1, \cdots, h_d$, asume for the moment that f is $N_d(\mu, \Sigma)$. Clever computations give $\int f^{1/2} dx = 2^{3d/4} \pi^{d/4} |\Sigma|^{1/4}$ and $\int |\ddot{f}_{ii}| dx = \sigma^{ii} 4e^{-1/2} / (2\pi)^{1/2}$. The leading terms for IMAD + IAB are of the form discussed above for both the GFP and the GFP-ASH, with $A_0 = (\frac{2}{\pi})^{1/2} B_0 \int f^{1/2} dx$, $A_i = B_i \int |\ddot{f}_{ii}| dx$. One arrives at

$$h_i^* = B_0^{2/(d+4)} (B_1 \cdots B_d)^{1/2(d+4)} B_i^{-1/2} K_d \, (|\Sigma| \sigma^{11} \cdots \sigma^{dd})^{1/2(d+4)} (\sigma^{ii})^{-1/2} n^{-1/(d+4)}, \tag{5.9}$$

where

$$K_d = 2^{3(d-4)/2(d+4)} \pi^{d/2(d+4)} e^{1/(d+4)}.$$

For the GFP, $B_0 = b_0^d$ and $B_i = \frac{1}{8}$, where $b_0 = \frac{1}{2} + \frac{1}{2\sqrt{2}} \log(1 + \sqrt{2})$. The cautious recommendation is therefore to estimate $\Sigma = (\sigma_{ij})$ in some robust way, and use

$$h_i^* = b_0^{2d/(d+4)} 2^{3d/2(d+4)} \pi^{d/2(d+4)} e^{1/(d+4)} \, (|\Sigma| \sigma^{11} \cdots \sigma^{dd})^{1/2(d+4)} (\sigma^{ii})^{-1/2} n^{-1/(d+4)}. \tag{5.10}$$

For example, $h^* = 1.551\sigma n^{-1/5}$ can be used as a starting value for h for the univariate frequency polygon. For the GFP-ASH, on the other hand, $B_0 = J(m_1)\cdots J(m_d)$ and $B_i = \frac{1}{12}\left(1+\frac{1}{2m_i^2}\right)$. Let us only give an answer for the case that $B_0 \doteq (\frac{2}{3})^{d/2}$ and $B_i \doteq \frac{1}{12}$ are good enough approximations (moderate m_i's will do). Then

$$h_i^* = 2^{(5d-4)/2(d+4)}3^{-(d-2)/(d+4)}\pi^{d/2(d+4)}e^{1/(d+4)}\,(|\Sigma|\sigma^{11}\cdots\sigma^{dd})^{1/2(d+4)}(\sigma^{ii})^{-1/2}n^{-1/(d+4)}. \tag{5.11}$$

In the one-dimensional case $h^* = 1.829\sigma n^{-1/5}$.

Similar reasoning can be used for the IMSE criterion. The typical IMSE has leading terms of the form

$$A_0(nh_1\cdots h_d)^{-1} + \sum_{i,j} A_{ij}h_i^2h_j^2,$$

see Theorems 1 and 3 and (5.2). Put $h_i = a_i n^{-\alpha}$, so that the IMSE becomes $A_0(a_1\cdots a_d)^{-1}$ $n^{-(1-d\alpha)} + \sum_{ij} A_{ij}a_i^2a_j^2n^{-4\alpha}$. The best choice is again $\alpha = 1/(d+4)$, giving

$$\text{IMSE } = \{A_0(a_1\cdots a_d)^{-1} + \sum_{i,j} A_{ij}a_i^2a_j^2\}n^{-4/(d+4)} + O(n^{-5/(d+4)}). \tag{5.12}$$

$a_1,\cdots,a_d$ remain to be specified. The values that minimize the expression in the brackets cannot be found in closed form solution (for $d \geq 3$), but can be found numerically for given values of $A_0, A_{11},\cdots,A_{dd}$. This requires the (first stage) estimation of the unknown quantities $\int(\ddot{f}_{ii})^2dx$, $\int \ddot{f}_{ii}\ddot{f}_{jj}dx$, by parametric or nonparametric methods. For example, if f is Gaussian with covariance matrix Σ, then $\int \ddot{f}_{ii}\ddot{f}_{jj}dx = (2\pi)^{-d/2}(\frac{1}{2})^{d/2}|\Sigma|^{-1/2}\{\frac{1}{2}(\sigma^{ij})^2 + \frac{1}{4}\sigma^{ii}\sigma^{jj}\}$, and this may be used to get at least starting values for $h_1,\cdots,h_d$. Comments about this are in Scott (1985a); here we shall only remark that this procedure, for the univariate frequency polygon, leads to $h^* = 2.153\sigma n^{-1/5}$, which can be compared with $h^* = 1.551\sigma n^{-1/5}$ obtained above with the L_1 view.

Scott (1985b) also discusses other methods of determining the smoothing parameters.

5.3. Concluding Remarks.

The previous subsection outlined how IMSE and IMAD + IAB expressions could be used to provide choices for smoothing parameters, i.e. window sizes for our generalizations of

histograms. In particular, it was seen that the natural IMAD + IAB criterion gave the more explicit recommendations, and also led to the reasonable measure (5.8) of how difficult the f at hand is to estimate. It is interesting to contrast these results with similar ones for the ordinary histogram (1.1), and for which we have already noted (5.4) and (5.5). Without going into the details, let us mark down that the best choice for $h_1, \cdots, h_d$ is

$$\tilde{h}_i = 2^{3/(d+2)}\pi^{-1/(d+2)}\Big(\int f^{1/2}dx\Big)^2\Big(\int |\dot{f}_1|dx \cdots \int |\dot{f}_d|dx\Big)^{1/(d+2)}\Big(\int |\dot{f}_i|dx\Big)^{-1/2}n^{-1/(d+2)} \tag{5.13}$$

based on the IMAD + IAB criterion, and that

$$\tilde{\Delta}(f) = \Big\{\Big(\int f^{1/2}dx\Big)^2 \int |\dot{f}_1|dx \cdots \int |\dot{f}_d|dx\Big\}^{1/(d+2)} \tag{5.14}$$

emerges as the reasonable measure of difficulty. If f is Gaussian with diagonal $\sigma_1^2, \cdots, \sigma_d^2$ covariance matrix, then

$$\tilde{h}_i = 2^{(3d+4)/(2d+4)}\pi^{d/(2d+4)}\sigma_i\, n^{-1/(d+2)}. \tag{5.15}$$

If the IMSE criterion is used instead, then

$$\tilde{h}_i = 6^{1/(d+2)}\Big\{\int (\dot{f}_1)^2dx \cdots \int (\dot{f}_d)^2dx\Big\}^{1/2(d+2)}\Big\{\int (\dot{f}_i)^2dx\Big\}^{-1/2}n^{-1/(d+2)} \tag{5.16}$$

is the best choice. For the Gaussian case

$$\begin{aligned}\tilde{h}_i &= 6^{1/(d+2)}(2\pi)^{d/(2d+4)}\sqrt{2}\{|\Sigma|\sigma^{11}\cdots\sigma^{dd}\}^{1/(2d+4)}(\sigma^{11})^{-1/2}n^{-1/(d+2)}\\ &\approx 3.50(|\Sigma|\sigma^{11}\cdots\sigma^{dd})^{1/(2d+4)}(\sigma^{ii})^{-1/2}n^{-1/(d+2)}.\end{aligned} \tag{5.17}$$

Is it sensible to choose window sizes and smoothing parameters on the basis of IMSE and IMAD + IAB? IMSE, for example, is really the expected loss ISE $= \int(\hat{f} - f)^2dx$. One can show that ISE/IMSE tends to one in probability, for all the estimators considered in this paper. This lends credibility to this criterion, and a similar justification can be given for IMAD + IAB. The rate at which ISE becomes close to IMSE may be slow, however; for example, one can show that $n^{d/(2d+4)}$(ISE/IMSE $-$ 1) has a limiting normal distribution in the histogram case, and that $n^{d/(2d+8)}$ (ISE/IMSE - 1) has a normal limit in the kernel estimator case.

Let us point out that it makes perfect sense to use for example the IMAD + IAB criterion over some specific region as a means of obtaining smoothing parameter values; the reasoning

of 5.2 can equally be applied. A more sophisticated procedure that, however, would complicate computational matters would be to use locally varying $h_1, \cdots, h_d$, say. These could, for example, be specified at a point x as the ones that give minimum IMSE (or estimated IMSE) in a ball of some fixed radius around x.

We have demonstrated that simple and computationally efficient variations on histograms can match for example kernel density estimators in performance. It is probably fair to point out, however, that both kernel estimators and the density estimators proposed in the present paper would have severe difficulties in being "statistically efficient" for anything but well behaved densities in higher dimensions, say for $d \geq 6$; they would require enormous sample sizes to detect possibly finer and interesting structure. One might turn to estimators based on projection pursuit methods, for example, to cope with such problems, see Huber (1985). One can hope, however, that methods like the GFP and the ASH can be useful as building blocks in such a more sophisticated set-up. Imagine, for example, that a "transformation pursuit" method was put to work on some six-dimensional data, and ended up giving a transformation from $(X_1, \cdots, X_6)$ to $(Y_1, \cdots, Y_6)$, say, having the property that (Y_1, Y_2, Y_3) and (Y_4, Y_5, Y_6) become practically independent. Then the GFP-ASH could be used to estimate the densities of (Y_1, Y_2, Y_3) and (Y_4, Y_5, Y_6), in a computationally and statistically efficient way. Then finally the density f for the original $(X_1, \cdots, X_6)$ is obtained by inverse transformation.

Appendix

Behind the statements displayed in Theorems 1, 2, 3 were a variety of Taylor expansions and approximations of integrals to Riemann sums. A more careful study of the remainder terms involved is called for now in order to actually prove the theorems.

Sometimes calculating the "next term" in an expansion, assuming the needed extra smoothness, is more informative than a proof, consisting as it must of bounding various remainder terms. This is done in a couple of instances below.

We shall have occasion to use a multivariate version of what Scott (1985a, p. 349) calls the generalized mean-value theorem. Let g be nonnegative and continuous on a cell $[a, b] = \prod_{i=1}^{d}[a_i, b_i]$. If φ is another continuous function on the cell, then

$$\int_{[a,b]} \varphi(x)g(x)dx = \varphi(x^*) \int_{[a,b]} g(x)dx \tag{A.1}$$

for some x^* in $[a, b]$. This can be proved as follows: It is trivially true if $\int_{[a,b]} g \; dx = 0$, so assume $g_0(x) = g(x)/\int_{[a,b]} g \; dx$ is a density on $[a, b]$. Then (A.1) amounts to $E_0 \; \varphi(X) = \varphi(x^*)$ where E_0 is expectation w.r.t. g_0. But φ carries the convex set $\prod_{i=1}^{d}[a_i, b_i]$ onto an interval, say $[c, d]$, and $E_0 \; \varphi(x)$ must be somewhere in that interval. (A.1) is also true for a nonpositive g but not necessarily for g taking both negative and positive values. (For example, $\int_{-1}^{1} x^2 \; dx \neq x^* \int_{-1}^{1} x \; dx$.)

A second fact to be used repeatedly below is given in the following lemma.

<u>Lemma A.1</u>. Assume $g : \mathbb{R}^d \to \mathbb{R}$ and its first order partial derivatives $g_1, \cdots, g_d$ are continuous and integrable. Then

$$\sum_k g(\xi_k)h_1 \cdots h_d = \int g \; dx + O\left(\sum_{i=1}^{d} h_i \int |g_i| dx\right), \tag{A.2}$$

where the sum is of over all cells, the union of which is $\mathbb{R}^d$, each cell has volume $h_1 \cdots h_d$, and ξ_k is an arbitrary point in cell number k.

An explicit and generous bound is available if the mixed higher order derivative $g_{1\cdots d}(x) = (\partial^d/\partial x_1 \cdots \partial x_d) \; g(x)$ exists, and all the functions $g_{ij}(x) = (\partial^2/\partial x_i \partial x_j)g(x)$, $i < j$; $g_{ijk}(x) =$

$(\partial^3/\partial x_i \partial x_j \partial x_k)g(x)$, $i < j < k; \cdots$, $g_{1\cdots d}(x)$ are integrable. In fact,

$$|\sum_k g(\xi_k)h_1 \cdots h_d - \int g\ dx| \le \sum_i h_i \int |g_i| dx + \sum_{i<j} h_i h_j \int |g_{ij}| dx$$
$$+ \sum_{i<j<k} h_i h_j h_k \int |g_{ijk}| dx + \cdots + h_1 \cdots h_d \int |g_{1\cdots d}| dx. \tag{A.3}$$

Proof: Consider a cell $\prod_{i=1}^d (a_i, b_i] = I$ and an arbitrary point $\xi = (\xi_1, \cdots, \xi_d)$ in it. Then

$$\begin{aligned} g(x) - g(\xi) &= \sum_{i=1}^d \{g(\xi_1, \cdots, \xi_{i-1}, x_i, x_{i+1}, \cdots, x_d) - g(\xi_1, \cdots, \xi_{i-1}, \xi_i, x_{i+1}, \cdots, x_d)\} \\ &= \sum_{i=1}^d \int_{\xi_i}^{x_i} g_i(\xi_1, \cdots, \xi_{i-1}, u_i, x_{i+1}, \cdots, x_d) du_i. \end{aligned}$$

Hence, using Fubini's theorem,

$$\begin{aligned} |\int_I g\ dx - g(\xi)h_1 \cdots h_d| &= |\int_I \{g(x) - g(\xi)\} dx| \\ &\le \sum_{i=1}^d \int_I \int_{a_i}^{b_i} |g_i(\xi_1, \cdots, \xi_{i-1}, u_i, x_{i+1}, \cdots x_d)| du_i\ dx \\ &= \sum_{i=1}^d (b_i - a_i) \int_I |g_i(\xi_1, \cdots, \xi_{i-1}, x_i, \cdots, x_d)| dx. \end{aligned}$$

Now use this bound for each cell:

$$\begin{aligned} |\int g\ dx - \sum_k g(\xi_k) h_1 \cdots h_d| &\le \sum_k |\int_{I_k} \{g(x) - g(\xi_k)\} dx| \\ &\le h_1 \int |g_1| dx + h_2 \sum_k \int_{I_k} |g_2(\xi_{k,1}, x_2, \cdots, x_d)| dx \\ &\quad + \cdots + h_d \sum_k \int_{I_k} |g_d(\xi_{k,1}, \cdots, \xi_{k,d-1}, x_d)| dx. \end{aligned}$$

That $\sum_k \int_{I_k} |g_i(\xi_{k,1}, \cdots, \xi_{k,i-1}, x_i, \cdots, x_d)| dx = \int |g_i| dx + \epsilon(h_1, \cdots, h_d)$, where $\epsilon(h_1, \cdots, h_d) \to 0$ as $\max\{h_1, \cdots, h_d\} \to 0$, follows from continuity and Riemann integrability of $|g_i(x)|$. This proves (A.2).

Next let us prove (A.3) under the stated extra assumptions. The proof is based on the

following monstrous algebraic decomposition:

$$
\begin{aligned}
g(x)-g(\xi) = & \sum_{i=1}^{d} \int_{\xi_i}^{x_i} g_i(x_1,\cdots,u_i,\cdots,x_d)du_i \\
& -\sum_{i<j} \int_{\xi_j}^{x_j}\int_{\xi_i}^{x_i} g_{ij}(x_1,\cdots,u_i,\cdots,u_j,\cdots,x_d)du_i\,du_j \\
& +\sum_{i<j<k} \int_{\xi_k}^{x_k}\int_{\xi_j}^{x_j}\int_{\xi_i}^{x_i} g_{ijk}(x_1,\cdots,u_i,\ldots,u_j,\cdots,u_k,\cdots,x_d)du_i\,du_j\,du_k \\
& +\cdots+(-1)^{d+1}\int_{\xi_d}^{x_d}\cdots\int_{\xi_1}^{x_1} g_{1\cdots d}(u_1,\cdots,u_d)du_1\cdots du_d.
\end{aligned}
\tag{A.4}
$$

For example, for $d=2$ (A.4) amounts to

$$
\begin{aligned}
g(x_1,x_2)-g(\xi_1,\xi_2) &= g(x_1,x_2)-g(\xi_1,x_2)+g(x_1,x_2)-g(x_1,\xi_2) \\
&\quad -\{g(x_1,x_2)-g(x_1,\xi_2)-g(\xi_1,x_2)+g(\xi_1,\xi_2)\}.
\end{aligned}
$$

Since a p-dimensional integral of $(\partial^p/\partial y_1\cdots\partial y_p)\ q(y_1,\cdots,y_p)$ over a rectangle can be written as an alternating sum of the 2^p corner values of q, (A.4) has 3^d-1 terms on the right hand side. (A.4) may be proved by carefully keeping track of the number of pluses and minuses in front of each particular term, and convince oneself that everything cancels except $g(x)-g(\xi)$.

Now (A.3) can be established. Consider the particular cell I first. Then

$$
\begin{aligned}
|\int_I\{g(x)-g(\xi)\}dx| \le & \sum_{i=1}^{d}\int_I\int_{a_i}^{b_i}|g_i(x_1,\cdots,u_i,\cdots,x_d)|du_i\,dx \\
& +\sum_{i<j}\int_I\int_{a_j}^{b_j}\int_{a_i}^{b_i}|g_{ij}(x_1,\cdots,u_i,\cdots,u_j,\cdots,x_d)|du_i\,du_j\,dx \\
& +\cdots+\int_I\int_{a_d}^{b_d}\cdots\int_{a_1}^{b_1}|g_{1\cdots d}(u)|du_1\cdots du_d\,dx \\
= & \sum_{i=1}^{d}(b_i-a_i)\int_I|g_i|dx+\sum_{i<j}(b_i-a_i)(b_j-a_j)\int_I|g_{ij}|dx \\
& +\cdots+(b_1-a_1)\cdots(b_d-a_d)\int_I|g_{1\cdots d}|dx.
\end{aligned}
$$

(A.3) follows by summing over all cells. ∎

<u>Remark A.1</u>. The lemma provides a multidimensional generalization of results reached by Freedman and Diaconis (1981), cf. their Corollary 2.24.

Still another Riemann sum lemma is needed. The lemma provides a strengthening of the previous one for the case that ξ_k, in the notation of (A.2) and (A.3), is the *centre point* in cell

k.

Lemma A.2. Let $g : \mathbb{R}^d \to \mathbb{R}$ have continuous and integrable partial derivatives g_i, g_{ij}, g_{ijs}. Let $(a_k - \frac{1}{2}h,\ a_k + \frac{1}{2}h] = \prod_{i=1}^d (a_{k,i} - \frac{1}{2}h_i,\ a_{k,i} + \frac{1}{2}h_i]$ be cell number k in a grid of cells with volume $h_1 \cdots h_d$. Then

$$\sum_k g(a_k) h_1 \cdots h_d = \int_{\mathbb{R}^d} g\ dx - \sum_{i=1}^d \frac{1}{24} h_i^2 \int_{\mathbb{R}^d} \ddot{g}_{ii}\ dx + O(h_1^3 + \cdots + h_d^3). \tag{A.5}$$

Proof: Consider a single cell first, and omit the subscript k for a moment. Then

$$\begin{aligned}
\int_{(a-\frac{1}{2}h,\ a+\frac{1}{2}h]} g(x)dx - g(a)h_1 \cdots h_d &= \int_{(a-\frac{1}{2}h,\ a+\frac{1}{2}h]} \{g(x) - g(a)\} dx \\
&= \int_{(a-\frac{1}{2}h,\ a+\frac{1}{2}h]} \left\{ \sum_{i=1}^d g_i(a)(x_i - a_i) + \frac{1}{2} \sum_{i,j} g_{ij}(\tilde{a}_x)(x_i - a_i)(x_j - a_j) \right\} dx \\
&= \frac{1}{2} \sum_{i=1}^d g_{ii}(\tilde{a}_i) \int_{(-\frac{1}{2}h,\ \frac{1}{2}h]} x_i^2 dx + \sum_{i<j} \int_{(a-\frac{1}{2}h,\ a+\frac{1}{2}h]} g_{ij}(\tilde{a}_x)(x_i - a_i)(x_j - a_j) dx,
\end{aligned}$$

where we used the generalized mean-value theorem for terms in the first sum. This is also possible for the second sum, but only after circumventive manoeuvring, which is necessary since $(x_i - a_i)(x_j - a_j)$ is neither nonnegative nor nonpositive on $(a - \frac{1}{2}h,\ a + \frac{1}{2}h]$.

Divide the cell into four regions, $\Omega_1, \cdot\cdot, \Omega_4$, where Ω_1 has $(a_i - \frac{1}{2}h_i,\ a_i)$ and $(a_j - \frac{1}{2}h_j,\ a_j)$ instead of $(a_i - \frac{1}{2}h_i,\ a_i + \frac{1}{2}h_i]$ and $(a_j - \frac{1}{2}h_j,\ a_j + \frac{1}{2}h_j]$, where Ω_2 similarly has $(a_i - \frac{1}{2}h_i,\ a_i)$ and $[a_j,\ a_j + \frac{1}{2}h_j]$, Ω_3 has $[a_i,\ a_i + \frac{1}{2}h_i]$ and $(a_j - \frac{1}{2}h_j,\ a_j)$, and finally Ω_4 has $[a_i,\ a_i + \frac{1}{2}h_i]$ and $[a_j,\ a_j + \frac{1}{2}h_j]$. The mean-value theorem can be employed for each of these four regions, and gives

$$\begin{aligned}
&\int_{(a-\frac{1}{2}h,\ a+\frac{1}{2}h]} \varphi(x)(x_i - a_i)(x_j - a_j)dx \\
&\qquad = \frac{1}{64} h_i h_j \{\varphi(\tilde{x}_1) - \varphi(\tilde{x}_2) - \varphi(\tilde{x}_3) + \varphi(\tilde{x}_4)\} h_1 \cdots h_d
\end{aligned}$$

for any continuous φ, for suitable $\tilde{x}_1$ in $\Omega_1, \cdot\cdot, \tilde{x}_4$ in Ω_4.

These observations lead to

$$\sum_k g(a_k)h_1\cdots h_d) - \int_{I\!R^d} g\,dx = \sum_k \int_{(a_k-\frac{1}{2}h,\,a_k+\frac{1}{2}h]} \{g(x)-g(a_k)\}dx$$

$$= -\sum_{i=1}^d \frac{1}{24}h_i^2 \sum_k g_{ii}(\tilde{a}_{k,i})h_1\cdots h_d$$

$$-\sum_{i<j}\frac{1}{64}h_ih_j\sum_k\{g_{ij}(\tilde{a}_{k,ij,1}) - g(\tilde{a}_{k,ij,2}) - g(\tilde{a}_{k,ij,3}) + g(\tilde{a}_{k,ij,4})\}h_1\cdots h_d,$$

where $\tilde{a}_{k,i}, \tilde{a}_{k,ij,1}, \cdots, \tilde{a}_{k,ij,4}$ all lie in cell number k. (A.5) follows upon application of Lemma A.1.

If g is only assumed to have second order continuous derivatives, then the reasoning above is still valid, but the remainder in (A.5) is then in general only $o(h_1^2+\cdots+h_d^2)$. ∎

Proof of Theorem 1: We will show

$$\int\{\text{bias}(x)\}^2dx = \sum_{i=1}^d \frac{49}{2880}h_i^4\int(\ddot{f}_{ii})^2dx$$
$$+\sum_{i<\ell}\frac{1}{32}h_i^2h_\ell^2\int\ddot{f}_{ii}\ddot{f}_{\ell\ell} + O\left(\sum_{i=1}^d h_i^5\right), \tag{A.6}$$

$$\int \text{Var}\{\hat{f}(x)\}dx = (\frac{2}{3})^d(nh_1\cdots h_d)^{-1} - \frac{1}{n}\int f^2dx + O\left(\frac{h_1^2+\cdots+h_d^2}{nh_1\cdots h_d}\right), \tag{A.7}$$

$$\int|\text{bias}(x)|dx \le \sum_{i=1}^d\frac{1}{8}h_i^2\int|\ddot{f}_{ii}|dx + O\left(\sum_{i=1}^d h_i^3\right), \tag{A.8}$$

$$\int \text{mad}\{\hat{f}(x)\}dx = (\frac{2}{\pi})^{1/2}\left(\frac{1}{2}+\frac{\log(1+\sqrt{2})}{2\sqrt{2}}\right)^d\int f^{1/2}dx\,(nh_1\cdots h_d)^{-1/2}$$
$$+O((nh_1\cdots h_d)^{-1} + \sum_{i=1}^d h_i(nh_1\cdots h_d)^{-1/2}), \tag{A.9}$$

which clearly suffices. Some of the error estimates can be improved, see Remark A.2 below.

Consider the Taylor expansions that led to (2.7) and (2.15). One has

$$f(x) = f(a) + \sum_{i=1}^d \dot{f}_i(a)(x_i-a_i) + \frac{1}{2}\sum_{i,\ell}\ddot{f}_{i\ell}(a)(x_i-a_i)(x_\ell-a_\ell)$$
$$+\frac{1}{6}\sum_{i,\ell,s}\dddot{f}_{i\ell s}(\tilde{a}_x)(x_i-a_i)(x_i-a_i)(x_s-a_s)$$

under the assumptions of the theorem, for some $\tilde{a}_x$ between a and x. Hence

$$\begin{aligned}
p_0(j) &= \int_{I_0(k;j)} f(x)dx \\
&= f(a)h_1\cdots h_d + \sum_{i=1}^{d} \dot{f}_i(a)\frac{1}{2}(-1)^{j_i+1}h_i(h_1\cdots h_d) \\
&+ \sum_{i=1}^{d} \ddot{f}_{ii}(a)\frac{1}{6}h_i^2(h_1\cdots h_d) + \sum_{i\neq \ell} \ddot{f}_{i\ell}(a)\frac{1}{8}(-1)^{j_i+j_\ell}h_ih_\ell(h_1\cdots h_d) \\
&+ \sum_{i,\ell,s} \frac{1}{6}\dddot{f}_{i\ell s}(\tilde{a}_{i\ell s,j}) \int_{[(j-1)h,\, jh]} x_ix_\ell x_s\, dx
\end{aligned}$$

by (A.1), where $\tilde{a}_{i\ell s,j}$ is somewhere in $[a+(j-1)h,\ a+jh]$. A more complete version of (2.7) and (2.15) is accordingly

$$\text{bias}(x) = \sum_{i=1}^{d} \ddot{f}_{ii}(a)\{\frac{1}{6}h_i^2 - \frac{1}{2}(x_i-a_i)^2\} + \delta(x), \tag{A.10}$$

where

$$\begin{aligned}
\delta(x) = &\sum_{j_1,\cdots,j_d} c_j(x)\sum_{i,\ell,s}\frac{1}{6}\dddot{f}_{i\ell s}(\tilde{a}_{i\ell s,j})\int_{[(j-1)h,\, jh]} x_ix_\ell x_s\, dx/(h_1\cdots h_d) \\
&-\sum_{i,\ell s}\frac{1}{6}\dddot{f}_{i\ell s}(\tilde{a}_x)(x_i-a_i)(x_\ell-a_\ell)(x_s-a_s).
\end{aligned} \tag{A.11}$$

Let us prove (A.6). Write (A.10) as

$$\text{bias}(x) = b(x) + \delta(x) = b(x) + \sum_{i,\ell,s}\delta_{i\ell s}(x).$$

The analysis of Section 2 implies together with the lemma above that

$$\begin{aligned}
\int\{b(x)\}^2dx &= \sum_{i=1}^{d}\frac{49}{2880}h_i^4\left\{\int(\ddot{f}_{ii})^2dx + O\left(\sum_{\ell=1}^{d}h_\ell\right)\right\} \\
&+\sum_{i<j}\frac{1}{32}h_i^2h_j^2\left\{\int \ddot{f}_{ii}\ddot{f}_{jj}\, dx + O\left(\sum_{\ell=1}^{d}h_\ell\right)\right\} \\
&= \sum_{i=1}^{d}\frac{49}{2880}h_i^4\int(\ddot{f}_{ii})^2dx + \sum_{i<j}\frac{1}{32}h_i^2h_j^2\int\ddot{f}_{ii}\ddot{f}_{jj}\, dx + O\left(\sum_{\ell=1}^{d}h_\ell^5\right).
\end{aligned} \tag{A.12}$$

(The lemma can be applied since $(\partial/\partial x_\ell)\ddot{f}_{ii}\ddot{f}_{jj} = \dddot{f}_{ii\ell}\ \ddot{f}_{jj} + \ddot{f}_{ii}\ \dddot{f}_{jj\ell}$ is absolutely integrable: $\int|\ \dddot{f}_{ii\ell}\ \ddot{f}_{jj}|dx \leq \{\int(\dddot{f}_{ii\ell})^2dx\}^{1/2}\{\int(\ddot{f}_{jj})^2dx\}^{1/2} < \infty$.) Next consider $\delta(x) = \sum_{i,\ell,s}\delta_{i\ell s}(x)$.

One has

$$\delta_{iii}(x) = \frac{1}{6}\left\{\sum_j c_j(x)\; \dddot{f}_{iii}\; (\tilde{a}_{iii,j})\frac{1}{4}(-1)^{j_i+1}h_i^3 - \dddot{f}_{iii}\; (\tilde{a}_x)(x_i - a_i)^3\right\},$$

so that

$$\int_{I(k)}\{\delta_{iii}(x)\}^2 \le \frac{2}{36}[\frac{1}{16}h_i^6\int_{I(k)}\{\sum_j c_j(x)\; \dddot{f}_{iii}\; (a_{iii,j})(-1)^{j_i+1}\}^2dx$$
$$+\int_{I(k)}\dddot{f}_{iii}\; (\tilde{a}_x)^2(x_i - a_i)^6dx].$$

Employing the generalized mean value theorem and some analysis we get the upper bound $\frac{1}{18\cdot16}(\frac{1}{3})^d h_i^6\{\sum_j \dddot{f}_{iii}\; (\tilde{a}_{iii,j})\}^2h_1\cdots h_d + \frac{1}{18\cdot7\cdot64}h_i^6\; \dddot{f}_{iii}\; (a_{iii}^*)^2h_1\cdots h_d$. Summing over all cells we arrive at $\int\{\delta_{iii}(x)\}^2dx = O(h_i^6)$. Similar analysis for the other terms, combined with the rough inequality $\delta(x)^2 \le 2^{d^3}\sum_{i,\ell,s}\delta_{i\ell s}(x)^2$, gives $\int\{\delta(x)\}^2dx = O\left(\sum_{\ell=1}^d h_\ell^6\right)$. This also shows, using $\int\{b(x)\}^2dx = O\left(\sum_{\ell=1}^d h_\ell^4\right)$, that

$$|\int b(x)\delta(x)dx| = O\left(\sum_{\ell=1}^d h_\ell^5\right).$$

All this proves (A.6).

(A.8) can be proved in a similar way; there are in fact fewer details to work through, and we omit them.

Next up is (A.7). An exact expression is

$$\mathrm{Var}\{\hat{f}(x)\} = (nh_1\cdots h_d)^{-1}\sum_j c_j(x)^2\frac{p_0(j)}{h_1\cdots h_d}$$
$$-\frac{1}{n}\left\{\sum_j c_j(x)\frac{p_0(j)}{h_1\cdots h_d}\right\}^2 .$$

for $x \in I(k) = (a - \frac{1}{2}h,\ a + \frac{1}{2}h]$, where

$$\begin{aligned}\frac{p_o(j)}{h_1\cdots h_d} &= f(a) + \sum_{i=1}^d \dot{f}_i(a)\frac{1}{2}h_i(-1)^{j_i+1}\\ &+\sum_{i=1}^d \ddot{f}_{ii}(\tilde{a}_{ii,j})\frac{1}{6}h_i^2 + \sum_{i\neq\ell}\ddot{f}_{i\ell}(\tilde{a}_{i\ell,j})\frac{1}{8}(-1)^{j_i+j_\ell}h_ih_\ell,\end{aligned} \tag{A.13}$$

with $\tilde{a}_{ii,j}$, $\tilde{a}_{i\ell,j}$ being somewhere in $(a+(j-1)h,\ a+jh]$. It is not difficult to get $\int_{I(k)} c_j(x)^2dx = (\frac{1}{3})^d h_1\cdots h_d$ and $\int_{I(k)} c_j(x)c_{j1}(x)dx = \prod_{i=1}^d(\frac{1}{3})^{I(j_i=j_i')}(\frac{1}{6})^{I(j_i\neq j_i')}(h_1\cdots h_d)$ for $j, j' \in \{0,1\}^d$,

from (2.9) and (2.10). It follows after some algebraic efforts that

$$\begin{aligned}\int_{I(k)} \operatorname{Var} \hat f(x)dx = (nh_1\cdots h_d)^{-1}\Big\{ &(\tfrac{2}{3})^d f(a)h_1\cdots h_d \\ &+\sum_{i=1}^{d} \frac{1}{6}h_i^2 \sum_j \ddot f_{ii}(\tilde a_{ii,j})(\tfrac{1}{3})^d h_1\cdots h_d \\ &+\sum_{i\neq \ell} \frac{1}{8}h_i h_\ell \sum_j (-1)^{j_i+j_\ell} \ddot f_{i\ell}(\tilde a_{i\ell,j})(\tfrac{1}{3})^d h_1\cdots h_d \Big\} \\ &-\frac{1}{n}\sum_{j,j'}\prod_{s=1}^{d}(\tfrac{1}{3})^{I(j_s=j_s')}(\tfrac{1}{6})^{I(j_s\neq j_s')}\Big[f(a)^2 \\ &+f(a)\sum_{i=1}^{d} \dot f_i(a)\frac{1}{2}h_i\{(-1)^{j_i+1}+(-1)^{j_i'+1}\}+\cdots\Big]h_1\cdots h_d.\end{aligned}$$

This implies, using the Riemann sum lemmas, that

$$\begin{aligned}\int_{\mathbb{R}^d} \operatorname{Var}\{\hat f(x)\}dx = (nh_1\cdots h_d)^{-1}&\left\{(\tfrac{2}{3})^d + O\left(\sum_{i=1}^{d} h_i^2\right)\right\} \\ &-\frac{1}{n}\left\{\int f^2dx + O\left(\sum_{i=1}^{d} h_i^2\right)\right\},\end{aligned}$$

i.e. (A.7) is true. (A minor technicality is that $\tilde a_j$ above may lie outside the cell $I(k)$; it is however at any rate in $(a-h,\ a+h]$, and a version of the Riemann sum lemma can be stated and proved for such occasions.)

Let us finally prove (A.9). Let $Y_0(k;j;i)$ be the indicator of the event that X_i falls in $I_0(k;j)=(a+(j-1)h,\ a+jh]$, so that $Y_0(k;j)=\sum_{i=1}^n Y_0(k;j;i)$. We may write

$$\hat f(x) - E\hat f(x) = (nh_1\cdots h_d)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_{n,i}(x)$$

where

$$A_{n,i}(x) = (h_1\cdots h_d)^{-1/2}\sum_{j_1,\cdots,j_d} c_j(x)\{Y_0(k;j;i)-p_0(j)\}.$$

Let

$$\begin{aligned}\sigma_n(x)^2 &= \operatorname{Var}\{A_{n,i}(x)\} \\ &= (h_1\cdots h_d)^{-1}\Big[\sum_j c_j(x)^2 p_0(j)\{1-p_0(j)\} - \sum_{j\neq j'} c_j(x)c_{j'}(x)p_0(j)p_0(j')\Big] \\ &= \sum_j c_j(x)^2\frac{p_0(j)}{h_1\cdots h_d} - h_1\cdots h_d\left\{\sum_j c_j(x)\frac{p_0(j)}{h_1\cdots h_d}\right\}^2.\end{aligned} \tag{A.14}$$

That

$$\left| |E\frac{1}{\sqrt{n}\sigma_n(x)}\sum_{i=1}^{n}A_{n,i}(x)| - (\frac{2}{\pi})^{1/2}\right| \leq B\frac{\rho_n(x)}{\sqrt{n}\sigma_n(x)^3},$$

where $\rho_n(x) = E|A_{n,i}(x)|^3$ and B is a universal constant, is a consequence of a nice Berry-Esséen type inequality proved in Devroye and Györfi (1985, Lemma 8, p. 90). Hence

$$E|\hat{f}_n(x) - f(x)| = (nh_1\cdots h_d)^{-1/2}\{(\frac{2}{\pi})^{1/2}\sigma_n(x) + \epsilon_n(x)\} \tag{A.15}$$

where $|\epsilon_n(x)| \leq B\ \rho_n(x)/\{\sqrt{n}\sigma_n(x)^2\}$.

Consider first the contribution from $\sigma_n(x)$. Combining (A.13) and (A.14) one sees that the leading terms in a Taylor expansion for $\sigma_n(x)^2$ is $f(a)\sum_j c_j(x)^2 + \sum_{i=1}^d \frac{1}{2}h_i\dot{f}_i(a)\sum_j(-1)^{j_i+1}$ $c_j(x)^2$. An expansion for $\sigma_n(x)$ accordingly starts out as

$$\begin{aligned}\sigma_n(x) &= f(a)^{1/2}\left\{\sum_j c_j(x)^2\right\}^{1/2} \\ &+\sum_{i=1}^{d}\frac{1}{2}h_i\ f(a)^{-1/2}\left\{\sum_j c_j(x)^2\right\}^{-1/2}\frac{1}{2}\dot{f}_i(a)\sum_j(-1)^{j_i+1}c_j(x)^2 + \cdots.\end{aligned} \tag{A.16}$$

This implies, after having properly tended to remainder terms, that

$$\begin{aligned}\int_{I(k)}\sigma_n(x)dx &= \{\frac{1}{2} + \frac{1}{2\sqrt{2}}\log(1+\sqrt{2})\}^d f(a)^{1/2}h_1\cdots h_d \\ &+O\left(\sum_{i=1}^{d}h_i\frac{|\dot{f}_i(a)|}{f(a)^{1/2}}h_1\cdots h_d\right),\end{aligned}$$

and consequently

$$\int_{\mathbb{R}^d}\sigma_n(x)dx = \{\frac{1}{2} + \frac{1}{2\sqrt{2}}\log(1+\sqrt{2})\}^d\int f^{1/2}dx + O\left(\sum_{i=1}^{d}h_i\right) \tag{A.17}$$

provided $\int|\dot{f}_i|/f^{1/2}dx$ is finite, $i = 1,\cdots,d$.

One will in fact usually have $O\left(\sum_{i=1}^d h_i^2\right)$ as an error term above. This is because the second term of (A.16) integrates to zero over $I(k)$ and because $\int f^{1/2}dx - \sum_k f(a_k)^{1/2}h_1\cdots h_d = \sum_{i=1}^d \frac{1}{24}h_i^2\int\frac{\partial^2}{\partial x_i^2}f(x)^{1/2}dx + O\left(\sum_{i=1}^d h_i^3\right)$. The error term of (A.17) can be shown to be $O\left(\sum_{i=1}^d h_i^2\right)$ provided $|\dot{f}_i\dot{f}_j|/f^{3/2}$ and $|\ddot{f}_{ij}|/f^{1/2}$ have finite integrals.

It remains to consider the contribution to IMAD from $\epsilon_n(x)$. The absolute third moment of $A_{n,i}(x)$ can be bounded as follows:

$$\rho_n(x) = (h_1 \cdots h_d)^{-3/2} \Big[\sum_j p_0(j) |c_j(x) - \sum_{j'} c_{j'}(x) p_0(j')|^3$$

$$\{1 - \sum_j p_0(j)\}| - \sum_j c_j(x) p_0(j)|^3 \Big]$$

$$\le (h_1 \cdots h_d)^{-3/2} \left\{ \sum_j p_0(j) c_{j'}(x) + \sum_j p_0(j) \sum_{j'} c_{j'}(x) p_0(j') + \sum_{j'} c_{j'}(x) p_0(j') \right\}$$

$$\le 3(h_1 \cdots h_d)^{-1/2} \sum_j c_j(x) \frac{p_o(j)}{h_1 \cdots h_d}.$$

(The bound can be somewhat improved, but not its order of magnitude.) By the bound quoted after (A.15), therefore,

$$|\epsilon_n(x)| \le \frac{3B}{(nh_1 \cdots h_d)^{1/2}} \frac{\sum_j c_j(x) p_0(j)/(h_1 \cdots h_d)}{\sigma_n(x)^2}.$$

Here $\sum_j c_j(x) p_0(j)/(h_1 \cdots h_d) = f(a) + O(\sum_{i=1}^d h_i)$ and $\sigma_n(x)^2 = f(a) \sum_j c_j(x)^2 + O(\sum_{i=1}^d h_i)$. It follows that

$$|\int_{I(k)} \epsilon_n(x) dx| \le \frac{3B}{(nh_1 \cdots h_d)^{1/2}} \int_{I(k)} \left[\left\{ \sum_j c_j(x)^2 \right\}^{-1} + O\left(\sum_{i=1}^d h_i \right) \right] dx$$

$$= (nh_1 \cdots h_d)^{-1/2} \left\{ 3(\frac{\pi}{2})^d B + O\left(\sum_{i=1}^d h_i \right) \right\} h_1 \cdots h_d.$$

The available bound on $\epsilon_n(x)$ therefore leads (only) to

$$|\int_R \epsilon_n(x) dx| \le 3(\frac{\pi}{2})^d B \text{ Volume}(R)(nh_1 \cdots h_d)^{-1/2} + O\left(\sum_{i=1}^d h_i/(nh_1 \cdots h_d)^{1/2} \right).$$

Combining this bound with (A.15) and (A.17) finally proves (A.9) and Theorem 1. ∎

Remark A.2. One sometimes gets a better idea of the typical magnitude of error terms if the "next term" of the expansion in question is computed, assuming extra smoothness. Consider for example the IMSE expression for the GFP obtained in Theorem 1. One may go through the calculations of Section 2 once more, but including the next order term in each

expansion. For example,

$$\begin{aligned}
\text{bias}(x) &\doteq \sum_{i=1}^{d} \ddot{f}_{ii}(a)\{\frac{1}{6}h_i^2 - \frac{1}{2}(x_i - a_i)^2\} \\
&+ \sum_{i=1}^{d} \dddot{f}_{iii}(a)\{\frac{1}{12}(x_i - a_i)h_i^2 - \frac{1}{6}(x_i - a_i)^3\} \\
&+ \sum_{i \neq \ell} \dddot{f}_{ii\ell}(a)\{\frac{1}{6}h_i^2(x_\ell - a_\ell) - \frac{3}{6}(x_i - a_i)^2(x_\ell - a_\ell)\}
\end{aligned}$$

is a more complete version of (2.15). And after a fair amount of algebraic and analytic details one arrives at

$$\begin{aligned}
\text{IMSE} &\doteq (\frac{2}{3})^d\{1 + \sum_{i=1}^{d} \frac{1}{8}h_i^2 \int \ddot{f}_{ii} dx\}(nh_1 \cdots h_d)^{-1} \\
&- \frac{1}{n}\left\{\int f^2 dx - \sum_{i=1}^{d} \frac{1}{4}h_i^2 \int (\dot{f}_i)^2 dx\right\} \\
&+ \sum_{i=1}^{d} \frac{49}{2880}h_i^4 \int (\ddot{f}_{ii})^2 dx + \sum_{i<\ell} \frac{1}{32}h_i^2h_\ell^2 \int \ddot{f}_{ii}\ddot{f}_{\ell\ell}\, dx \\
&+ \sum_{i=1}^{d} \frac{71}{241920}h_i^6 \int (\dddot{f}_{iii})^2 dx \\
&+ \sum_{i \neq \ell}\{\frac{49}{34560}h_i^4h_\ell^2 \int (\dddot{f}_{ii\ell})^2 dx + \frac{7}{5760}h_i^4h_\ell^2 \int \dddot{f}_{iii}\dddot{f}_{i\ell\ell}\, dx\} \\
&+ \sum_{\substack{i,j,\ell \\ \text{distinct}}} \frac{1}{768}h_i^2h_j^2h_\ell^2 \int \dddot{f}_{ii\ell}\dddot{f}_{jj\ell}\, dx.
\end{aligned}$$

<u>Proof of Theorem 2</u>. The steps to be taken are similar to, but sometimes more involved, than the ones displayed in the proof of Theorem 1, and most of the details will be omitted.

One can give a longer and exact expression for the bias than the Taylor approximation (3.7), using an exact version of (3.5). Then one is led to an analogue of the result stated before (3.8). A key observation is then that

$$\sum_k \dot{f}_j(x_k)^2\delta_1 \cdots \delta_d = \int (\dot{f}_j)^2 dx + O\left(\sum_{i=1}^{d} \delta_i^3\right),$$

which follows by Lemma A.2, since $\frac{\partial^2}{\partial x_\ell^2}\{\dot{f}_j(x)\}^2 = 2\{\ddot{f}_{j\ell}(x)\}^2 + 2\dot{f}_j(x)\,\dddot{f}_{j\ell\ell}(x)$ integrates to zero. Hence the error term corresponding to summing up $\sum_{j=1}^{d} \dot{f}_j(x_0)^2 \frac{1}{12}\delta_j^2\ \delta_1 \cdots \delta_d$ terms is

$O(\sum_{i=1}^d \delta_i^5) = O(\sum_{i=1}^d h_i^5/m_i^5)$. But summing up $\ddot{f}_{jj}(x_0)^2 \delta_j^4\, m_j^4\, \delta_1 \cdots \delta_d$ terms, for example, is seen to lead to remainder with magnitude $O(\sum_{i=1}^d \delta_i^5\, m_i^4) = O(\sum_{i=1}^d h_i^5/m_i)$, and similarly for the other contributions, so $O(\sum_{i=1}^d h_i^5/m_i)$ is the remainder for the integrated squared bias.

One can similarly scrutinize the contributions to IAB, using Lemmas A.1 and A.2. There are remainders of size $O(\sum_{i=1}^d \delta_i^3) = O(\sum_{i=1}^d h_i^3/m_i^3)$ and $O(\sum_{i=1}^d \delta_i^3\, m_i^2) = O(\sum_{i=1}^d h_i^3/m_i)$, and the latter one dominates.

Next consider IVAR. An exact version of (3.5) based on an exact second order Taylor expansion is

$$
\begin{aligned}
p(i_1, \cdots, i_d)/(\delta_1 \cdots \delta_d) &= f(x_0) + \sum_{i=1}^{d} \dot{f}_j(x_0) i_j \delta_j \\
&+ \frac{1}{2} \sum_{j=1}^{d} \ddot{f}_{jj}(\tilde{x}_{i,jj})(i_j^2 + \frac{1}{12})\delta_j^2 + \sum_{j<\ell} \ddot{f}_{j\ell}(\tilde{x}_{i,j\ell}) i_j i_\ell\, \delta_j \delta_\ell,
\end{aligned}
$$

where $\tilde{x}_{i,jj}$ and $\tilde{x}_{i,j\ell}$ are in $(x_0 + (i - \frac{1}{2})\delta,\ x_0 + (i + \frac{1}{2}\delta]$; $1 - m_j \le i_j \le m_j - 1$, $j = 1, \cdots, d$. (We ignore some slight complications of no consequence that enter the situation for $\sum_{j<\ell}$ terms when one or more i_j is zero; then $(x_j - x_{0,j})(x_\ell - x_{0,\ell})$ is neither nonnegative nor nonpositive on the cell, and the generalized mean-value theorem must be applied with cutting and pasting. See the proof of Lemma A.2.) From a formula in Section 3.2, therefore,

$$
\begin{aligned}
\mathrm{Var}\{\hat{f}_{\mathrm{ASH}}(x)\} = (nh_1 \cdots h_d)^{-1} \{ (\tfrac{2}{3})^d \prod_{j=1}^{d} (1 + \frac{1}{2m_j^2}) f(x_0) \\
+ \frac{1}{2} \sum_{j=1}^{d} \delta_j^2 \sum_{i} \ddot{f}_{jj}(\tilde{x}_{i,jj})(i_j^2 + \frac{1}{12}) \frac{w(i)^2}{m_1 \cdots m_d} + \cdots \} \\
- \frac{1}{n} \{ f(x_0) + \frac{1}{2} \sum_{j=1}^{d} \delta_j^2 \sum_{i} \ddot{f}_{jj}(\tilde{x}_{i,jj})(i_j^2 + \frac{1}{12}) \frac{w(i)}{m_1 \cdots m_d} + \cdots \}^2.
\end{aligned}
$$

Now

$$
\delta_j^2 \sum_{i} \ddot{f}_{jj}(\tilde{x}_{i,jj})(i_j^2 + \frac{1}{12}) \frac{w(i)^2}{m_1 \cdots m_d} = O(\delta_j^2\, m_j^2 |\ddot{f}_{jj}(x_0^*)|),
$$

where $|f_{jj}(x_0^*)| = \max_{x \in [x_0 - h,\ x_0 + h]} |\ddot{f}_{jj}(x)|$. Even though x_0^* may be outside of $C_0 = (x_0 - \frac{1}{2}\delta,\ x_0 + \frac{1}{2}\delta]$, one may show that the sum of $\delta_j^2\, m_j^2 |\ddot{f}_{jj}(x_0^*)| \delta_1 \cdots \delta_d$, over all cells, is $\delta_j^2 m_j^2 \{ \int |\ddot{f}_{jj}| dx + O(h_1 + \cdots + h_d) \}$, using techniques as in Lemmas A.1 and A.2. The result

from all this is that

$$\int \operatorname{Var}\{\hat{f}_{\text{ASH}}(x)\}dx = (nh_1 \cdots h_d)^{-1}\Big\{(\tfrac{2}{3})^d \prod_{j=1}^{d}\left(1+\frac{1}{2m_j^2}\right) + O\left(\sum_{j=1}^{d} \delta_j^2\right)$$
$$+ O\left(\sum_{j=1}^{d} h_j^2\right)\Big\} - \frac{1}{n}\Big\{\int f^2 dx + O\left(\sum_{j=1}^{d} h_j^2\right)\Big\}.$$

It remains only to take care of IMAD. One has

$$\int_{C_0} \operatorname{mad}\{\hat{f}_{\text{ASH}}(x)\}dx = (n\delta_1 \cdots \delta_d)^{-1/2} E\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^{n} B_{n,k}\Big|$$

by (3.14), writing

$$B_{n,k} = (\delta_1 \cdots \delta_d)^{1/2} \sum_{i_1,\cdots,i_d} \frac{w(i_1,\cdots,i_d)}{m_1 \cdots m_d}\{Y(i_1,\cdots i_d;\ k) - p(i_1,\cdots,i_d)\},$$

where $Y(i_1,\ldots,i_d;\ k) = Y(i;k)$ is an indicator for the event that X_k falls in $C_0(i) = (x_0 + (i-\frac{1}{2})\delta,\ x_0 + (i+\frac{1}{2})\delta]$, so that $Y(i) = \sum_{k=1}^{n} Y(i;k)$. Using Lemma 8 of Devroye and Györfi (1985, p. 90) again,

$$E\Big|\frac{1}{\sqrt{n}}\sum_{k=1}^{n} B_{n,k}\Big| = (\frac{2}{\pi})^{1/2}\sigma_n + \epsilon_n,$$

say, where $\sigma_n^2 = \operatorname{Var} B_{n,k}$ and $|\epsilon_n| \leq B\ \rho_n/(\sqrt{n}\sigma_n^2)$, $\rho_n = E|B_{n,k}|^3$. Going once more through arguments that resemble those used in the proof of Theorem 1, the result is that the contribution to IMAD from the σ_n part is

$$(nh_1 \cdots h_d)^{-1/2}\left\{(\frac{2}{\pi})^{1/2}(\frac{2}{3})^{d/2}\prod_{j=1}^{d}\left(1+\frac{1}{2m_j^2}\right)^{1/2}\int f^{1/2}dx + O\left(\sum_{j=1}^{d} h_j^2\right)\right\}$$

provided the integrals $\int |\ddot{f}_{ij}|/f^{1/2}dx$ are finite, and that the contribution from the ϵ_n part over some bounded region R is less than

$$(n\delta_1 \cdots \delta_d)^{-1/2} 3B\ n^{-1/2} m_1 \cdots m_d\ (\delta_1 \cdots \delta_d)^{-1/2} \operatorname{Volume}(R)$$
$$= 3B\frac{m_1^2 \cdots m_d^2}{nh_1 \cdots h_d} \operatorname{Volume}(R).$$

This proves Theorem 2. ∎

<u>Proof of Theorem 3</u>: We are in a position to use techniques displayed in the proofs of Theorems 1 and 2 and omit most details.

One has

$$E\hat{f}_{\text{GFP-ASH}}(x) = \sum_{\ell_1,\cdots,\ell_d} \frac{T(\ell_1,\cdots,\ell_d)}{m_1\cdots m_d}\frac{p(\ell_1,\cdots,\ell_d)}{\delta_1\cdots\delta_d},$$

and one can write down an exact exprssion for $p(\ell_1,\cdots,\ell_d)/(\delta_1\cdots\delta_d)$ based on an exact third order Taylor expansion for f. (4.7) displays the first terms in the expression for $p(\ell_1,\cdots,\ell_d)/(\delta_1\cdots\delta_d)$. The next term, i.e. the exact remainder, has $\sum_{i=1}^d \frac{1}{6}\,\dddot{f}_{iii}\,(\tilde{a}_{iii,\ell})(\ell_i^3 - \frac{3}{2}\ell_i^2 + \ell_i - \frac{1}{4})\delta_i^3$ for some $\tilde{a}_{iii,\ell}$ in $(a+(\ell-1)\delta,\ a+\ell\delta]$, and some other terms of the same magnitude. Reasoning as in the proof of Theorem 1, one has

$$\text{bias}(x) = b(x) + \delta(x) = b(x) + \sum_{i,j,k}\delta_{ijk}(x),$$

with $b(x) = \sum_{i=1}^d \ddot{f}_{ii}(a)\{\frac{1}{12}(m_i^2+1)\delta_i^2 - \frac{1}{2}(x_i-a_i)^2\}$ for $x \in (a-\frac{1}{2}\delta,\ a+\frac{1}{2}\delta]$, and, for example, $\delta_{iii}(x) = \frac{1}{6}\delta_i^3\sum_\ell \dddot{f}_{iii}\,(\tilde{a}_{iii,\ell})\frac{T(\ell)}{m_1\cdots m_d}(\ell_i^3 - \frac{3}{2}\ell_i^2 + \ell_i - \frac{1}{4})$ for x in the same cell. These facts can be used to show that $\int b(x)^2dx$ equals the right hand side of (4.9) with remainder $O\left(\sum_{i=1}^d h_i^4\delta_i\right) = O\left(\sum_{i=1}^d h_i^5/m_i\right) = O\left(\sum_{i=1}^d \delta_i^5 m_i^4\right)$. Also, $\int \delta_{iii}(x)^2dx = O\left(\sum_{i=1}^d \delta_i^6 m_i^4\right) = O\left(\sum_{i=1}^d h_i^6/m_i^2\right)$, and $|\int b(x)\delta(x)dx| = \{O\left(\sum_{i=1}^d h_i^4\right) O\left(\sum_{i=1}^d h_i^6/m_i^2\right)\}^{1/2} = O\left(\sum_{i=1}^d h_i^5/m_i\right)$. ∎

Similar analysis shows

$$\text{IAB} \le \sum_{i=1}^d \frac{1}{12}h_i^2\left(1+\frac{1}{2m_i^2}\right)\int |\ddot{f}_{ii}|dx + O\left(\sum_{i=1}^d h_i^3/m_i\right).$$

Next up is IVAR. One has

$$\begin{aligned}\text{Var}\{\hat{f}_{\text{GFP-ASH}}(x)\} = (nh_1\cdots h_d)^{-1}&\sum_\ell \frac{T(\ell)^2}{m_1\cdots m_d}\frac{p(\ell)}{\delta_1\cdots\delta_d}\\ &-\frac{1}{n}\{\sum_\ell \frac{T(\ell)}{m_1\cdots m_d}\frac{p(\ell)}{\delta_1\cdots\delta_d}\}^2,\end{aligned}$$

and the expansion for $p(\ell)/(\delta_1\cdots\delta_d)$ mentioned above can be used to get

$$\begin{aligned}\int_{(a-\frac{1}{2}\delta,\ a+\frac{1}{2}\delta]}\{\hat{f}_{\text{GFP-ASH}}(x)\}dx &= (nh_1\cdots h_d)^{-1}\{(\frac{2}{3})^d f(a)\delta_1\cdots\delta_d\\ &\quad+\sum_{i=1}^d O(\delta_i^2 m_i^2 \ddot{f}_{ii}(a))\delta_1\cdots\delta_d\}\\ &\quad-\frac{1}{n}\{f(a)^2\delta_1\cdots\delta_d + \sum_{i=1}^d O(\delta_i^2 m_i^2 f(a)\ddot{f}_{ii}(a))\delta_1\cdots\delta_d\},\end{aligned}$$

and hence, using once more the Riemann sum lemmas,

$$\text{IVAR} = (nh_1 \cdots h_d)^{-1}\{(\frac{2}{3})^d + O\left(\sum_{i=1}^{d} h_i^2\right)\} - \frac{1}{n}\{\int f^2 dx + O\left(\sum_{i=1}^{d} h_i^2\right)\}.$$

It remains to deal with IMAD. One has

$$\text{mad}\{\hat{f}_{\text{GFP-ASH}}(x)\} = (nh_1 \cdots h_d)^{-1/2} E|\frac{1}{\sqrt{n}} \sum_{k=1}^{n} C_{n,k}(x)|$$

from Section 4.3, where

$$C_{n,k}(x) = (h_1 \cdots h_d)^{-1/2} \sum_{\ell} T(\ell)\{Y(\ell;k) - p(\ell)\}.$$

Once more

$$E|\frac{1}{\sqrt{n}} \sum_{k=1}^{n} C_{n,k}(x)| = (\frac{2}{\pi})^{1/2}\tau_n(x) + \eta_n(x),$$

where $\tau_n(x)^2 = \text{Var}\, C_{n,k}(x)$ and $|\eta_n(x)| \le B\ E|C_{n,k}(x)|^3/\{\sqrt{n}\tau_n(x)^2\}$. Reasoning as at previous occasions one may show

$$|\int_R \eta_n(x)dx| \le \frac{\text{const.}}{(nh_1 \cdots h_d)^{1/2}}$$

for each bounded region R, with the const. in question proportional to the volume of R. Also,

$$\begin{aligned}
\tau_n(x)^2 &= \sum_{\ell} \frac{T(\ell)^2}{m_1 \cdots m_d} \frac{p(\ell)}{\delta_1 \cdots \delta_d} - h_1 \cdots h_d \left\{\sum_{\ell} \frac{T(\ell)}{m_1 \cdots m_d} \frac{p(\ell)}{\delta_1 \cdots \delta_d}\right\}^2 \\
&= f(a) \prod_{j=1}^{d} \frac{2m_j^2 + 1 - 6u_j(1-u_j)}{3m_j^2} + \sum_{i=1}^{d} \delta_i \dot{f}_i(a) \sum_{\ell} (\ell_i - \frac{1}{2}) \frac{T(\ell)^2}{m_1 \cdots m_d} \\
&\quad + \sum_{i=1}^{d} \frac{1}{2} \delta_i^2 \ddot{f}_{ii}(a) \sum_{\ell} (\ell_i^2 - \ell_i + \frac{1}{3}) \frac{T(\ell)^2}{m_1 \cdots m_d} \\
&\quad + \sum_{i \ne j} \frac{1}{2} \delta_i \delta_j \ddot{f}_{ij}(a) \sum_{\ell} (\ell_i - \frac{1}{2})(\ell_j - \frac{1}{2}) \frac{T(\ell)^2}{m_1 \cdots m_d} \\
&\quad + O\left(\sum_{i=1}^{d} \delta_i^3\right),
\end{aligned}$$

so that

$$\tau_n(x) = f(a)^{1/2} \prod_{j=1}^{d} \left(\frac{2m_j^2 + 1 - 6u_j(1 - u_j)}{3m_j^2} \right)^{1/2}$$

$$+ \sum_{i=1}^{d} \frac{1}{2} \delta_i f(a)^{-1/2} \prod_{j=1}^{d} \left(\frac{2m_j^2 + 1 - 6u_j(1 - u_j)}{3m_j^2} \right)^{-1/2} \dot{f}_i(a) \frac{(2m_i^2 + 1)(u_i - \frac{1}{2})}{2m_i^2 + 1 - 6u_i(1 - u_i)}$$

$$+ \sum_{i=1}^{d} O\Big(\delta_i^2 m_i^2 \{|\ddot{f}_{ii}(a)| f(a)^{-1/2} + \dot{f}_i(a)^2 f(a)^{-3/2}\}\Big).$$

It follows that

$$\int_{(a-\frac{1}{2}\delta,\, a+\frac{1}{2}\delta]} \tau_n(x)dx = f(a)^{1/2} \delta_1 \cdots \delta_d \, J(m_1) \cdots J(m_d)$$

$$+ \sum_{i=1}^{d} O\Big(\delta_i^2 m_i^2 \{|\ddot{f}_{ii}(a)| f(a)^{-1/2} + \dot{f}_i(a)^2 f(a)^{-3/2}\}\Big) \delta_1 \cdots \delta_d$$

and finally that

$$\int \tau_n(x)dx = J(m_1) \cdots J(m_d) \int f^{1/2} dx + O\left(\sum_{i=1}^{d} h_i^2 \right),$$

provided $|\ddot{f}_{ij}|/f^{1/2}$ and $|\dot{f}_i \dot{f}_j|/f^{3/2}$ have finite integrals.

This proves Theorem 3. ∎

Acknowledgements.

This paper has been written during my stay at Stanford University's Department of Statistics with grants from the Norwegian Computing Center and the Royal Norwegian Council for Scientific and Industrial Research. I am grateful to David Scott for many inspiring discussions. Thanks are also due to Linda Davis for discussions and to Judi Davis for efficient typing, without which a critical deadline would not have been met.

References

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.

Epanechnikov, V. A. (1969). Nonparametric estimation of a multi-dimensional probability density. *Theory of Probab. Appl.* **14**, 153–159.

Freedman, D. A. and Diaconis, P. (1981). On the histogram as a density estimator. *Z. Wahrsch. verw. Gebiete* **57**, 453–476.

Huber, P. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13**, 435–525.

Jones, M. C. and Lotwick, H. W. (1984). A remark on Algorithm AS 176: Kernel density estimation using the Fast Fourier transform. *Appl. Statist.* **33**, 120–122.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.

Scott, D. W. (1985a). Frequency polygons: Theory and applications. *J. Amer. Statist. Assoc.* **80**, 348–354.

Scott, D. W. (1985b). Average shifted histograms: Effective nonparametric density estimation in several dimensions. *Ann. Statist.* **13**, 1024–1040.

Scott, D. W. and Thompson, J. R. (1983). Probability density estimation in higher dimensions. In *Computer Science and Statistics: Proc. Fifteenth Symposium on the Interface* (J. E. Gentle, ed.), 173–179.

Silverman, B. W. (1982). Algorithm AS 176: Kernel density estimation using the Fast Fourier transform. *Appl. Statist.* **31**, 93–99.

Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.* **7**, 328–340.